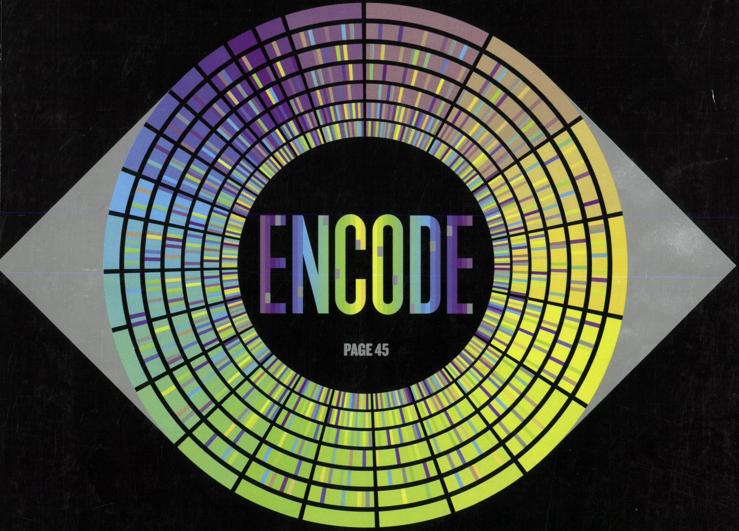
nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



# ENCODE

PAGE 45

## GUIDEBOOK TO THE HUMAN GENOME

*The ENCODE project in print and online*

**PLANETARY SCIENCE**

### LAST RAYS OF THE SUN

*Venerable Voyager 1 can still surprise*

PAGES 20 & 124

**PALAEONTOLOGY**

### HARNESSING FOSSIL POWER

*How China's feathered dinosaurs sparked revolution*

PAGE 22

**TOXICOLOGY**

### RETHINK ON RISK DATA

*Why the EPA should acknowledge uncertainty*

PAGE 27

➲ NATURE.COM/NATURE

6 September 2012 £10

Vol. 489, No. 7414

36>

9 770028 083095

Research Experience
ducation
Medicinal Chemistry and Biochemistry
Medicinal Chemistry, Upper Second Class

naturejobs.com

**Upload your CV at *Naturejobs* and let's take your career sky high**

Ready to take your job search to the next level? Upload your CV and cover letter at the new improved naturejobs.com and take full advantage of the world's largest science jobs board. Your saved information will be immediately available, so you can quickly and easily apply for one of the **10,000+ science vacancies** found at naturejobs.com.

**To find out more, visit:**
**www.naturejobs.com/cvupload**

**Follow us on:**

nature publishing group npg

Introducing the BD FACSJazz™
Perfectly tuned cell sorting from BD Biosciences.

BD FACSJazz
BD
Helping all people live healthy lives

Breathtaking solo performance.

The BD FACSJazz™ system begins a new era in cell sorting with stellar performance, benchtop fit, and an affordable price.

The BD FACSJazz can be configured with up to 3 lasers and 8 parameters to support your application needs—offloading sorting demand at core labs or meeting the needs of individuals.

Requiring less than 2 x 2 ft (20 x 20 in, 51 x 51 cm) of bench space, the BD FACSJazz also addresses biosafety concerns with an optional custom designed biosafety cabinet that meets personnel and product protection standards for both a Class II Type A2 biosafety cabinet and the National Sanitation Foundation International Standard 49.

In addition to factory-optimized settings, the BD FACSJazz comes standard with BD FACS™ Sortware sorter software to simplify use. This innovative software is specifically designed for cell sorting and features comprehensive control for acquisition, sorting, and analysis.

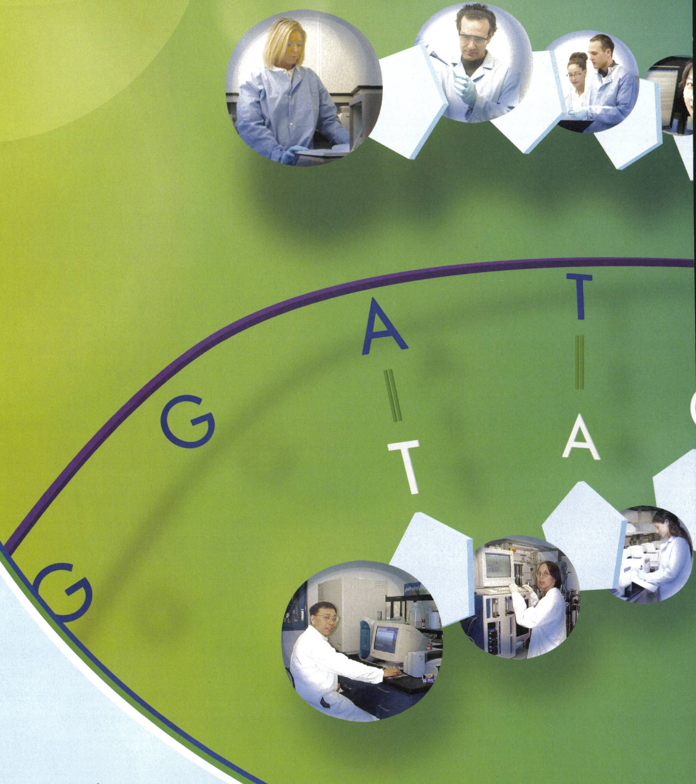
For legendary results, just add your creativity. Learn how at **bdbiosciences.com/go/jazz**

Class 1 Laser Product.
For Research Use Only. Not for use in diagnostic or therapeutic procedures.
BD, BD Logo and all other trademarks are property of Becton, Dickinson and Company. © 2012 BD
23-14201-00

**BD Biosciences**
2350 Qume Drive
San Jose, CA 95131
bdbiosciences.com

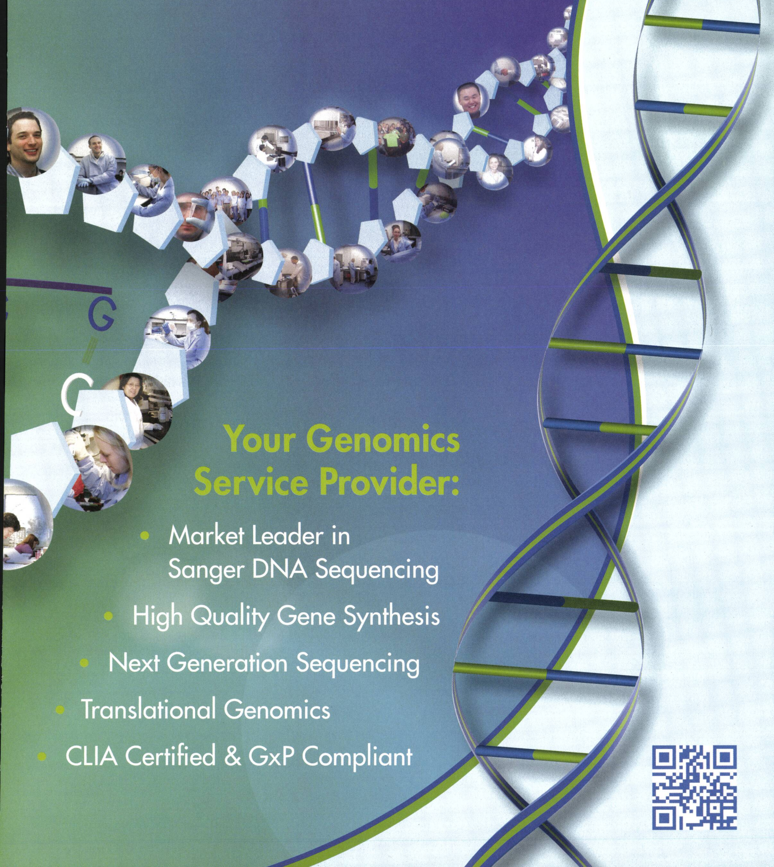
GENEWIZ®
Solid science. Superior service.

Contributing to th



www.genewiz.com

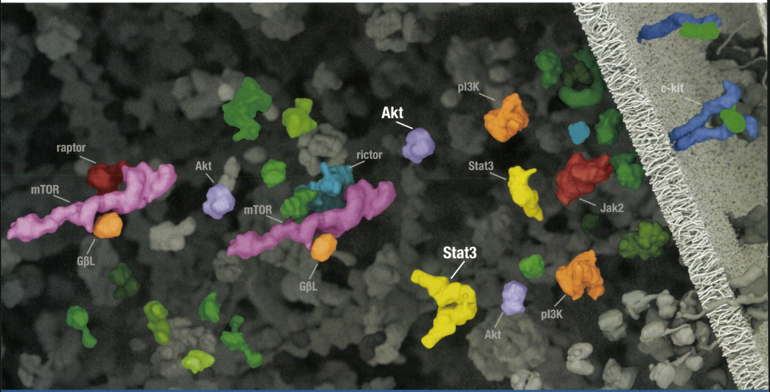
dvancement of Global Life Sciences
Specializing in Genomics Since 1999



Your Genomics Service Provider:

- Market Leader in Sanger DNA Sequencing
- High Quality Gene Synthesis
- Next Generation Sequencing
- Translational Genomics
- CLIA Certified & GxP Compliant

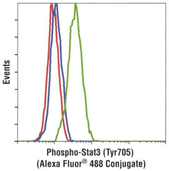
Antibodies and Related Reagents for Signal Transduction Research



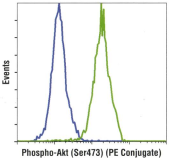
# Conjugated Antibodies for Flow Cytometry *from Cell Signaling Technology*

## Unparalleled product quality, validation, and technical support

- :: The highest quality antibodies provide you with the brightest signal and the lowest background.
- :: Extensive validation by our in-house flow cytometry group means optimization is not left up to you, the user.
- :: Technical support provided by our flow cytometry specialists translates into a thorough, fast, and accurate response.
- :: High quality custom conjugations of our off-the-shelf antibodies are also available, please inquire about sizing and pricing options.
- :: Multiple fluorochrome options are available to meet your unique multiplexing needs.



**Phospho-Stat3 (Tyr705) (D3A7) XP® Rabbit mAb (Alexa Fluor® 488 Conjugate) #4323:** Flow cytometric analysis of Jurkat cells, untreated (blue) or IFN-α treated (green), using #4323 compared to nonspecific control antibody (red).



**Phospho-Akt (Ser473) (D9E) XP® Rabbit mAb (PE Conjugate) #5315:** Flow cytometric analysis of Jurkat cells, untreated (green) or treated with LY294002 (PI3 Kinase Inhibitor) #9901, Wortmannin (PI3 Kinase Inhibitor) #9551, and U0126 (MEK1/2 Inhibitor) #9903 (blue), using #5315.

for quality products you can trust...

www.cellsignal.com

Cell Signaling
TECHNOLOGY®

**Orders** (toll-free) 1-877-616-2355 | **Technical support** (toll-free) 1-877-678-8324 support@cellsignal.com | **Inquiries** info@cellsignal.com | **Environmental Commitment** eco.cellsignal.com

**For Research Use Only. Not For Use In Diagnostic Procedures.**

©2012 Cell Signaling Technology, Inc. / XP® and Cell Signaling Technology® are trademarks of Cell Signaling Technology, Inc.
Alexa Fluor® is a trademark of Molecular Probes, Inc.

# CONTENTS

6 September 2012 / Vol 489 / Issue No 7414

## nature | ENCODE

### EDITORIAL

5 **Spinning threads**

### INTRODUCTION

45 A feast of data

### FEATURE

46 **The human encyclopaedia**

### COMMENT

49 **Lessons for big-data projects**
*Ewan Birney*

### NEWS & VIEWS

52 **ENCODE explained**
*Joseph R Ecker; Wendy A Bickmore; Inês Barroso; Jonathan K Pritchard &Yoav Gilad; Eran Segal*

### RESEARCH

57 **An integrated encyclopedia of DNA elements in the human genome**
*The ENCODE Project Consortium*

75 **The accessible chromatin landscape of the human genome**
*R E Thurman* et al.

83 **An expansive human regulatory lexicon encoded in transcription factor footprints**
*S Neph* et al.

91 **Architecture of the human regulatory network derived from ENCODE data**
*M B Gerstein* et al.

101 **Landscape of transcription in human cells**
*S Djebali* et al.

109 **The long-range interaction landscape of gene promoters**
*A Sanyal, B R Lajoie, G Jain & J Dekker*

### CAREERS

165 **A single cog in a complex machine**

nature.com/encode

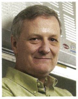
COVER ILLUSTRATION: CARL DETORRES

## THIS WEEK

### EDITORIALS

5 RESEARCH
**Accountable and transparent**
The NIH should ensure that there is public disclosure of conflicts of interest

6 ASTROPHYSICS
**Moonlight drive**
The road goes ever on and on for the Voyager probes



### WORLD VIEW

7 **We must be open about our mistakes**
*Jim Woodgett*
Transparency is the key to a healthy scientific enterprise

### RESEARCH HIGHLIGHTS

8 SELECTIONS FROM THE SCIENTIFIC LITERATURE
Convoluted hydrocarbons / An early digging mammal / Methane at sea / An exoplanet dies

### SEVEN DAYS

10 THE NEWS IN BRIEF
Extinction threat for invertebrates / Schizophrenia drug fails trial / Australia joins Europe's Emissions Trading System / Funding boost for arXiv

## NEWS IN FOCUS

13 MEDICAL RESEARCH
Drug developers regroup after Alzheimer's failures

14 CONSERVATION
India's assessment of forest health called into question

16 PUBLIC HEALTH
Change to patent rules sparks protest

17 CHEMISTRY
Row puts focus on reporting standards

19 INFORMATICS
Biological databases face funding cuts

20 ASTROPHYSICS
Surprises at the edge of the Solar System

### FEATURE

22 PALAEONTOLOGY
**The ground breaker**
Xing Xu is revolutionizing ideas about dinosaur evolution

## CAREERS

167 CAREER BRIEFS

NATUREJOBS ADVERTISING FEATURE
Career Expo Preview / Spotlight on Wallonia

## COMMENT

27 TOXICOLOGY
**Rethink chemical risk assessments**
The EPA needs to speed up its decisions
*George M Gray & Joshua T Cohen*

### BOOKS & ARTS

29 ARCHITECTURE
**Life in stone**
*Georgina Ferry*

30 TECHNOLOGY
**Baroque geekery**
*Tim Boon*

31 BOOKS IN BRIEF

32 Q&A
**Maths demystifier**
Glen Whitney's MoMath museum, open in December, aims to inspire

### CORRESPONDENCE

33 NASA priorities / Saving tigers / Stem-cell regulation / Siting wind farms

### OBITUARY

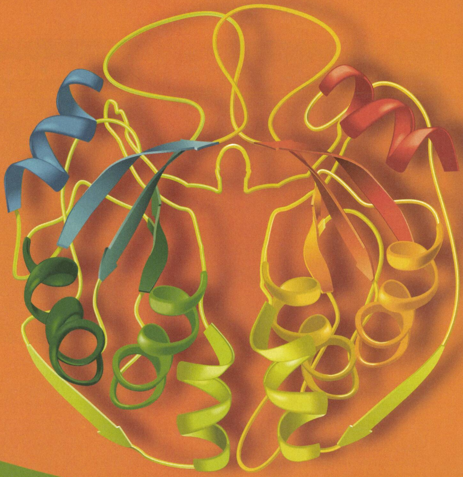
34 **Martin Fleischmann (1927–2012)**
*Philip Ball*

### FUTURES

170 **If only ...**
*Tony Ballantyne*

# 7,000 Human Proteins Made from Human Cells.

More Assays!
Better Assays!



## The Largest Collection of Mammalian Cell-Derived Proteins

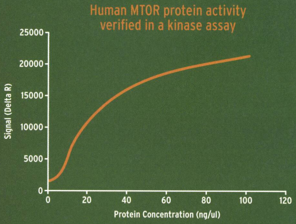
**Protein functional studies**

**Compound screening**

**Immunogens for antibody production**

**Assay standards: ELISA, Mass Spec, and more**

**Bulk production available**



Human MTOR protein activity verified in a kinase assay

ORIGENE
Your Gene Company

origene.com/protein

# CONTENTS

6 September 2012 / Vol 489 / Issue No 7414

## RESEARCH

### NEW ONLINE

35 Papers published this week at nature.com

### NEWS & VIEWS

36 MATERIALS SCIENCE
**A hard concept in soft matter**
A highly fracture-resistant hydrogel
*Kenneth R Shull* **SEE LETTER P.133**

37 COSMOLOGY
**The lithium problem**
Detection of interstellar lithium beyond the Milky Way
*Garik Israelian* **SEE LETTER P.121**

38 NEUROSCIENCE
**Lessons from heartbreak**
Dopamine neurons control pheromone sensitivity and mating in *Drosophila*
*Aki Ejima* **SEE LETTER P.145**

39 CLIMATE CHANGE
**Brief but warm Antarctic summer**
Long record of climate change on the Antarctic Peninsula
*Eric J Steig* **SEE LETTER P.141**

41 SURFACE SCIENCE
**Separation by reconfiguration**
Water-responsive membranes that separate emulsions
*Robert W Field*

42 ASTRONOMY
**Outflows from the first quasars**
Gas ejection from an accreting supermassive black hole
*Daniel Mortlock*

43 STRUCTURAL BIOLOGY
**A protein engagement RING**
The elusive association between E2–ubiquitin and RING E3 enzymes
*Christopher D Lima & Brenda A Schulman*
**SEE ARTICLE P.115**

52 FORUM: GENOMICS
**ENCODE explained**
The impact of cataloguing the human genome's functional elements
*Joseph R Ecker; Wendy A Bickmore; Inês Barroso; Jonathan K Pritchard & Yoav Gilad; Eran Segal* **SEE ENCODE PAGES**

### ARTICLES

57 ENCODE **An integrated encyclopedia of DNA elements in the human genome**
*The ENCODE Project Consortium*

75 ENCODE **The accessible chromatin landscape of the human genome**
*R E Thurman* et al.

83 ENCODE **An expansive human regulatory lexicon encoded in transcription factor footprints**
*S Neph* et al.

91 ENCODE **Architecture of the human regulatory network derived from ENCODE data**
*M B Gerstein* et al.

101 ENCODE **Landscape of transcription in human cells**
*S Djebali* et al.

115 STRUCTURAL BIOLOGY **Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis**
*A Plechanovová, E G Jaffray, M H Tatham, J H Naismith & R T Hay* **SEE N&V P.43**

### LETTERS

109 ENCODE **The long-range interaction landscape of gene promoters**
*A Sanyal, B R Lajoie, G Jain & J Dekker*

121 ASTRONOMY **Observation of interstellar lithium in the low-metallicity Small Magellanic Cloud**
*J C Howk, N Lehner, B D Fields & G J Mathews* **SEE N&V P.37**

124 ASTRONOMY **No meridional plasma flow in the heliosheath transition region**
*R B Decker, S M Krimigis, E C Roelof & M E Hill*

128 TECHNOLOGY **Flexible metal-oxide devices made by room-temperature photochemical activation of sol–gel films**
*Y-H Kim* et al.

133 MATERIALS SCIENCE **Highly stretchable and tough hydrogels**
*J-Y Sun* et al. **SEE N&V P.36**

137 EARTH SCIENCE **Activation of old carbon by erosion of coastal and subsea permafrost in Arctic Siberia**
*J E Vonk* et al.

141 CLIMATE **Recent Antarctic Peninsula warming relative to Holocene climate and ice-shelf history**
*R Mulvaney* et al. **SEE N&V P.39**

145 NEUROSCIENCE **Dopamine neurons modulate pheromone responses in *Drosophila* courtship learning**
*K Keleman* et al. **SEE N&V P.38**

150 STEM CELLS **Neuronal circuitry mechanism regulating adult quiescent neural stem-cell fate decision**
*J Song* et al.

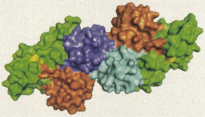
155 CANCER **Heterodimeric JAK–STAT activation as a mechanism of persistence to JAK2 inhibitor therapy**
*P Koppikar* et al.

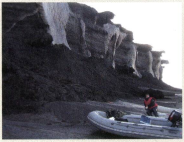
160 IMMUNOLOGY **Endogenous antigen tunes the responsiveness of naive B cells but not T cells**
*J Zikherman, R Parameswaran & A Weiss*

STRUCTURAL BIOLOGY

**Close call**

How E2–ubiquitin and RING E3 enzymes tie the knot. **PAGE 115**





EARTH SCIENCE

**C change**

Thawing Arctic permafrost could release many tons of carbon. **PAGE 137**

# Ten Truths About Conservation Agriculture and Smallholder Farmers

BY HOWARD G. BUFFETT

*In the July 5, 2012 issue we provided examples on what we have learned during our 20-year experience working in conservation agriculture in both large and small systems. Now we want to share with you what others have learned.*

There are growing numbers of smallholder farmers throughout the developing world who are successfully implementing conservation agriculture practices adapted to specific local conditions and existing crop and livestock production customs. As a result, farmers are using a wide variety of reduced tillage techniques and various means to protect soils with organic ground covers. It is clear conservation agriculture, like all agriculture practices, offers no single "silver-bullet solution" that satisfies every farm condition. The constraints and availability of natural resources, local climatic conditions, socio-economic policies and other factors all have a role in shaping a farmer's approach. However, conservation agriculture offers unique and critically-needed solutions to many of the challenges all farmers face, requiring us to make every effort to develop its full potential.

A rigorous scientific understanding of the success factors involved in conservation agriculture for smallholder farmers is only now being established. Consequently, much of the evidence that demonstrates conservation agriculture's potential for this group of farmers is drawn from a limited number of sub-scale development initiatives and underfunded research by scientific institutions and civil society organizations.

We do not have all of the answers but there is no question we need more public and private sector support for research in this area. There are many promising signs that investing in conservation agriculture practices will lead to improved farmer livelihoods, increased food security and enhanced local and global environmental quality. Examples of how it is being successfully implemented, and the scientific research that is providing solutions to key issues, are discussed in the following overview of conservation agriculture's basic truths and their applicability to smallholder farmers.

## Truth #1: Smallholder farmers who adopt integrated conservation practices can realize a higher return on investment in terms of labor savings, net income and improved soil quality.

- The Food and Agriculture Organization of the United Nations (FAO) studied nearly 5,000 smallholder farmers who adopted conservation agriculture practices in four different regions of Tanzania and Kenya following their participation in farmer field schools. Farmers who adopted conservation agriculture substantially reduced their labor inputs while also improving their crop yields from 26%-100% or more over a period of three to ten years. Farmers who used appropriate direct seeding equipment (e.g. a manual 'jabber') could plant a field of 0.4 hectares in three to four hours as compared to conventional tillage where three people working with hand hoes needed an entire day. (Shetto, et al. "Conservation Agriculture as Practiced in Tanzania: Three Case Studies." 2007.)

- With conservation agriculture's reduced labor and time requirements for planting new crops, farmers are better able to sow seeds during the optimal planting 'window.' Studies indicate that for each day that seeding is delayed past the optimal planting period, harvested crop yields can be reduced by 1-1.5%. (Olaf Erenstein, "Zero Tillage in the Rice-Wheat Systems of the Indo-Gangetic Plains." IFPRI Discussion Paper # 00916. International Food and Policy Research Institute. Washington D.C. 2009.)

- With the encouragement and support of the Ministry of Agriculture and several non-governmental organizations (NGOs), many smallholder farmers in Ghana have replaced their reliance on traditional 'slash and burn' cultivation methods with no-till and crop residue mulching practices that include the use of herbicides to control weeds. The International Maize and Wheat Improvement Center (CIMMYT) and the Crop Research Institute conducted field surveys of farmers in 30 different villages to determine the relative labor requirements of the two systems. These surveys found that the overall family labor inputs were 27% lower for those farmers who adopted conservation agriculture and their maize crop yields were 57% greater than those achieved by farmers who continued to rely on slash and burn practices. (J. Ekboir, et al. "Impact of No-Till Technologies in Ghana." CIMMYT Economics Program Paper 02-01. 2002.)

- The World Food Programme's (WFP) Purchase for Progress (P4P) program has reported that participating Nicaraguan smallholder farmers' adoption of conservation agriculture practices has enabled them to reduce their production costs by 40%. (Ken Davies, WFP/P4P Coordinator. 2012.)

- The carbon sequestration potential of conservation agriculture practices is variable. However, many studies have shown that soil organic carbon (SOC, aka carbon) levels under no-till are much higher in the first 10 centimeters of topsoil than in soils under conventional tillage (e.g. approximately 75% higher at a five cm depth and 40% higher at a 10 cm depth). SOC levels are similar at 20-30 cm depths for both systems. Models indicate that total SOC under conservation agriculture could be 10-30% greater than for conventional tillage. (Robert, et al. Soil Carbon Sequestration for Improved Land Management. FAO. 2001.)

- A meta analysis of 67 long-term agriculture experiments that compared soil carbon sequestration rates between no-till and conventional tillage practices found that the transition from tillage to no-till practices could continually sequester an average 480 kg C/ha/year over a period of 15-20 years until the SOC levels reach a stable equilibrium. (T. West and W. Post. "Soil Organic Carbon Sequestration Rates by Tillage and Crop rotation: A Global Data Analysis." Soil Science Society of America Journal. 66. 2002.)

- Ratan Lal, the Director of the Carbon Management and Sequestration Center at Ohio State University has reported that the SOC levels of natural ecosystems have been significantly depleted by historical land misuse and poor management of soils. Lal notes that if farmers adopted "recommended management practices" that include conservation agriculture with no-till farming, residue mulching, cover cropping, crop rotations, appropriate use of both organic and inorganic fertilizers and other related land stewardship techniques, approximately 100 to 1000 kg C/ha/year could be sustainably sequestered until a new equilibrium level of SOC is achieved over a 25-50 year period. (Lal. "Beyond Copenhagen: mitigating climate change and achieving food security through soil carbon sequestration." Food Security. Springer. 2010; and M. Jarecki and R. Lal. "Carbon Management for Soil Carbon Sequestration." Critical Reviews in Plant Sciences/ Vol. 22. 2003.)

### Truth #2: A combination of education and site-specific analysis will help balance competing uses for crop residue.

- An International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) study of the dual use (soil conservation vs. livestock feed) of crop residues in Niger indicated that the nutrient content of some crop residues is unevenly distributed throughout the plant segments (e.g. stalks, stems, panicles, etc.). In such crop residues the lower structural biomass has very little nutritious value as livestock feed. It is believed that for some crops (e.g. millet), nearly 40% of total residues could be more valuable in contributing to overall farm productivity and farmer income if used as ground cover for soil conservation that increase crop yields rather than as fodder for livestock. (Powell and Fussell, "Nutrient and Structural Carbohydrate Partitioning in Pearl Millet." Agronomy Journal. Vol. 85, July, 1993.)

- The Consultative Group on International Agricultural Research (CGIAR) Systemwide Livestock Programme's analysis of economic trade-offs between using crop residues for ground cover mulches or for fodder in West Africa found that optimum returns were possible with varying levels of crop residue retention across different locations and crop varieties. In some cases the majority of residues should be used for soil cover and nutrient recycling; in other cases using a majority of residues for fodder provided farmers with the highest return. (IITA et al. "Balancing livestock needs and soil conservation: Final Report." 2011.)

- In some communities, farmers protect their crop residue mulches from free grazing livestock by discussing the importance of organic ground cover for crop productivity with village governing organizations and their neighbors. In some areas community self-governance groups prohibit free grazing livestock practices, resulting in residues being retained for mulch. (CA-SARD Project. CA Cases from Tanzania.)

- Farmers have cultivated livestock fodder crops on field contours and set aside areas and then harvested these crops as 'cut and carry' feed for corralled livestock. These practices have resulted in crop residues being available for use as mulch and enabled farmers to effectively collect manure for use as organic fertilizer. (Bolliger et al. "Taking Stock of the Brazilian Zero-Till Revolution: A Review of Landmark Research and Farmers' Practice." Elsevier. 2006.)

### Truth #3: Diverse varieties of cover crops, crop rotations and substantial residue retention can reduce reliance on herbicides.

- As farmers gain experience in growing cover crops that control weeds, weed pressures can be reduced. Ground cover residues and green manure crops can also suppress weed infestations in later years to a level that enables farmers to use a fraction of the herbicides that were originally needed. (Steiner, et al. "Weed Management in Conservation Tillage Systems." African Conservation Tillage Network. Information Series #8.; and Kofi Bofa. Ghana. personal communication. 2012.)

A farmer in Ghana collects and spreads leaves and crop roots onto his soil to increase organic matter. When I asked him why he was doing this, he replied, "to give my soil life." All farmers understand the importance of healthy soil.

- In long-term field trials in Malawi, CIMMYT researchers found that the use of cover crops and green manures over a period of seven cropping seasons could control weeds without the need for further applications of herbicides. (Bram Govaerts. CIMMYT. personal communication. 2012.)

- Multi-year crop rotations with different plant varieties have significantly reduced the need for herbicides to control weeds. A study in Iowa of Low External Input (LEI) farming practices compared conventional two-year corn/soybean rotations with three- and four-year rotations of corn with N-fixing crops. The study found that over a four-year period the longer rotations significantly reduced the need for herbicides (i.e. by 76% and 82% respectively) and synthetic fertilizers (i.e. by 59% and 74% respectively). The rotational crops' allelopathic biochemical and ground cover competition properties may explain the superior weed control results. (M. Liebman, et al. "Agronomic and Economic Performance Characteristics of Conventional and Low-External Input Cropping Systems in the Central Corn Belt." Agronomy Journal. Vol. 100. 2008.)

### Truth #4: Efforts to identify crops adapted to local conditions, building capacity of farm organizations and improving smallholder access to key inputs are gaining traction.

- Since 2006, the Program for Africa's Seed Systems (PASS) has provided technical and financial support to local research institutions for the development of their institutional capacity and scientific breeding of high-yielding, disease resistant maize, cassava and rice varieties. PASS also links research institutions with local seed companies and provides direct loans and equity investments to emerging seed companies where necessary. These efforts ensure farmers are aware of improved seeds, can access them and get the extension services they need to improve yields. To date over 60 independent seed companies have been created (with another 40 planned by 2017) and over 9,000 agro-dealers have been trained and networked, resulting in 373,283 MT of seed sold to African farmers. In 2010, 21 MSc and 10 PhD students supported by PASS graduated, helping to build critical technical and research capacity. Our foundation has supported the expansion of PASS's efforts into Sierra Leone, Liberia and South Sudan. (AGRA Annual Report, 2010.)

- NGOs and agricultural research institutions are developing and propagating green manure cover crops and N-fixing plants and woody legumes that are adapted to African agroclimatic conditions. In addition, Farmer Field Schools are encouraging farmers to establish local seed banks in which a variety of rotational and cover crop seeds and seedlings are produced and exchanged within the community. (Eotuleo Farmer Field School: CA-SARD Project. CA Cases from Tanzania.)

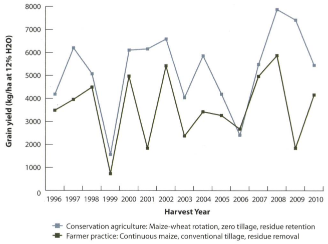
- The World Food Programme's P4P is a 21-country pilot project to source commodities for WFP's food assistance efforts from local smallholder farmers. As part of the program in Central America, P4P provides conservation agriculture extension services to participating farmer organizations. One element of this training introduced N-fixing bean varieties to women farmers in Guatemala. By adding these legumes to their intercropping and crop rotation practices, the farmers improved the nutrition of their diets, introduced an additional source of income from the sale of these beans and naturally produced N fertilizer for their fields. (WFP. P4P January Update 2011.)

- CARE International's Hillside Conservation Agriculture Project (HICAP) in Tanzania encouraged participating farmers to cultivate N-fixing cover crops (e.g. pigeon pea, lablab and cowpeas) that could be harvested for sale as starter seeds to other farmers. The HICAP project also facilitated seed distribution to farmers in the area to support Farmer Field Schools and demonstration plots. (CARE HICAP Project Annual Report to HGBF. 2012.)

### Truth #5: Many smallholders who adopt conservation agriculture experience increased crop yields.

- Long-term conservation agriculture field research in Mexico demonstrates that conservation agriculture practices for maize and wheat production consistently achieved higher yields over conventional practices from the first year onward. In the initial year of the study conservation agriculture maize and wheat yields were respectively 20% and 33% greater. Conservation agriculture yields for both crops were particularly better during drought years when conservation agriculture's enhanced soil moisture retention properties had a positive impact on grain production. (Erenstein, et al. "Adapting No-tillage Agriculture to the Conditions of Smallholder Maize and Wheat Farmers in the Tropics and sub-Tropics." No-Till Farming Systems. Eds. T. Goddard, et al. Bangkok: World Association of Soil and Water Conservation. 2008.)

Effect of Conservation Agriculture on Maize Yields in El Batan, Mexico, 1996-2010



Adapted from: CIMMYT analysis, Verhulst, N., Nelissen, V., Jespers, N., Haven, H., Sayre, K.D., Raes, D., Deckers, J., Govaerts, B., 2011.

- Immediate crop yield improvements were achieved by farmers participating in CARE International's farmer field school demonstration project in Tanzania. These farmers had an average yield gain of 87% over the conventional control plot yields in the first year. (CARE International HICAP Project report to HGBF, 2012.)

### Truth #6: Conservation agriculture's focus on soil health, reduced erosion and an emphasis on crop diversity reduces the exposure of smallholders to crop failures.

- Long term field research by CIMMYT indicates that zero tillage with retention of residues significantly improves soil moisture levels and enables crop yields that are far greater than those from conventional practices during extended dry periods. (Verhulst, et al. "Conservation agriculture as a means to mitigate and adapt to climate change..." 2011.)

- The Tropical Soil Biology and Fertility Institute of The International Center for Tropical Agriculture (CIAT-TSBF) led a multi-institutional research study of mixed crop-livestock systems in Mozambique that evaluated smallholder intercropping of maize with pigeonpea and cowpea legumes. This intercropping practice provided farmers with improved surface cover of their fields, valuable N-fixing soil inputs, increased resilience to low rainfall conditions and better food security. (Rusinamhodzi, et al. "Productivity of Maize-Legume Intercropping under No-Till in Central Mozambique." Regional Conservation Agriculture Symposium. Johannesburg, South Africa. 2011.)

- Ghanaian farmers surveyed by CIMMYT affirmed that they benefited from reduced risks as a result of their adoption of conservation agriculture. (Ekboir, et al. "Impact of No-Till Technologies in Ghana." CIMMYT. Economics Program Paper 02-01. 2002.)

### Truth #7: Investing in conservation agriculture requires a long-term commitment to achieve widespread adoption.

In Mozambique we discovered farmers using a homemade roller crimper to control weeds, protect seedlings from the heat and "provide food for their soil," as one of the farmers told me.

- In Brazil, conservation agriculture is a well-established practice, but it took a tremendous commitment first by early adopters, and then by government and extension agents to encourage widespread adoption. Local land user clubs were invaluable to changing farmers' mindsets by promoting the concept and helping each other in their conservation agriculture adoption efforts. The most innovative conservation agriculture farmers continually share their knowledge of best practices with their neighbors. They also share modifications to direct seeding equipment adapted to local soil and crop residue conditions. (Bolliger et al. "Taking Stock of the Brazilian Zero-Till Revolution: A Review of Landmark Research and Farmers' Practice." Elsevier. 2006; and Ribeiro, et al. "Comprehensive inventory and assessment of existing knowledge on sustainable agriculture in Latin America." CIRAD. 2007.)

- Catholic Relief Services' (CRS) conservation agriculture development project in Zimbabwe has been a multi-year effort to introduce farmers to conservation agriculture practices and support their adoption. An initial 650 farmers adopted conservation agriculture practices in the first year (2004/2005) and 54,416 farmers were practicing conservation agriculture at the end of five years. (Mutsindikwa, et al. "Conservation Agriculture "a winners' choice."" CRS-Zimbabwe. FAO. CA Regional Workshop. 2011.)

### Truth #8: Plants use synthetic fertilizer more efficiently when combined with conservation agriculture practices.

- A team of soil scientists assessed the efficiencies of improved crop yield gains with inorganic fertilizer inputs when applied to soils with higher levels of organic carbon. The study team concluded that "for efficient nutrient utilization, inorganic fertilizer must be combined with organic matter, water harvesting, conservation agriculture and controlling soil erosion in site-specific integrated soil fertility management strategies." Their recommendations clearly identify the important role of conservation agriculture practices in enabling smallholder farmers to maximize the benefits of their use of synthetic fertilizer inputs. (Bationo, et al. "African Soils: Their Productivity and Profitability of Fertilizer Use." African Fertilizer Summit. Nigeria. 2006.)

- In the absence of conservation agriculture, if fertilizer use is stopped (because of lack of availability or affordability), production levels will likely fall, soil quality will not have been improved, and farmers will still lack the skills needed to sustainably build soil health to achieve long-term productivity. (Dorward, et al. "Towards 'smart' subsidies in agriculture? Lessons from recent experience in Malawi." Natural Resources Perspectives 116. Overseas Development Institute. September 2008; and Jayne, et al. "Fertilizer Subsidies in Eastern and Southern Africa." COMESA African Agricultural Markets Programme. Policy Synthesis #2. 2009.)

### Truth #9: Investment in farmer organizations helps farmers develop the skills and confidence to encourage and support the adoption of new practices like conservation agriculture.

- Cooperative groups of local farmers convincingly demonstrate best practices and serve as mentors to their peers. They are also better able to finance, purchase and share tools, exchange improved seeds and agree to follow livestock management programs that protect residues and cover crops for uses that improve soil health and crop yields. Where local communities have been technically and organizationally developed, the eventual discontinuation of external assistance has not led to widespread farmer rejection of productive conservation agriculture practices. (Penning de Vries, editor. "Bright Spots Demonstrate Community Success in African Agriculture." IMWI Working Paper 102. 2005.)

- It is critical to focus on enhancing farmer skills, knowledge and opportunities to directly participate in research and testing of techniques and technologies that best leverage local resources. By building farmers' individual and collective capacities to manage integrated conservation agriculture practices, their skills and experiences mature and they will become less reliant on remotely sourced inputs and provisions of external funding. (Agriculture at a Crossroads. IAASTD. Island Press. 2009.)

### Truth #10: Smallholder farmers themselves must participate as partners in the research, development and demonstration efforts to advance conservation agriculture adoption.

- In the early 1990s, most maize and wheat farmers in the Yaqui Valley of Mexico were burning their crop residues. With the technical assistance of CIMMYT and other agricultural development organizations' field demonstrations of raised bed maize cultivation methods, improved irrigation techniques and mentoring support, in the span of one decade nearly all farmers had changed their practices and now retain their residues for soil conservation purposes. (Verhulst, et al. "Conservation agriculture as a means to mitigate and adapt to climate change, a case study from Mexico." Designing Agricultural mitigation for smallholders in developing countries. Wollenberg L. editor. Earthscan. 2011.)

- In the span of forty years, Brazil has undergone a remarkable transition from relying on traditional and conventional tillage practices to today's condition where it is a major agricultural producer with 75% of cropland under no-till or reduced tillage. While this transformation substantially relied on innovating and expanding large scale farming operations, it also included adoption by hundreds of thousands of smallholder farmers. This farming revolution was accomplished with strong and persistent support from government and private sector investment in developing techniques, technologies and cropping practices to promote soil fertility and control erosion while producing profitable agricultural outputs. These efforts included focused research programs, development and commercialization of no-till technologies, extensive field training and other initiatives. (Bolliger, et al. "Taking Stock of the Brazilian 'Zero-Till Revolution.' " Elsevier. 2006.)

With long-term investment, and a commitment to research, innovation, and developing a support system to encourage the adoption of conservation agriculture, it is clear the developing world can transform its approach to farming to be more productive and more sustainable for current and future generations.

---

THE HOWARD G.
BUFFETT
FOUNDATION

### About The Howard G. Buffett Foundation

Established in 1999, The Howard G. Buffett Foundation's primary mission is to improve the standard of living and quality of life for the world's most impoverished and marginalized populations. The Foundation's focus is on international programs that operate in challenging environments, including conflict and post-conflict countries. The Foundation has supported more than 120 agricultural projects in over 40 countries and more than 50 nutrition projects in over 30 countries. The Foundation believes achieving global food security requires all countries—including the United States—to adapt its agricultural practices and policies to meet long-term agricultural needs and successfully address hunger and malnutrition. To learn more, visit www.thehowardgbuffettfoundation.org.

# nature

6 September 2012 / Vol 489 / Issue No. 7414

## EDITORIAL

**LONDON, MUNICH, WASHINGTON DC, NEW YORK, BOSTON, PARIS, SAN FRANCISCO, TOKYO** **nature@nature.com**

**EDITOR-IN-CHIEF:** Philip Campbell
**EXECUTIVE EDITOR:** Nick Campbell
**PUBLISHING EXECUTIVE EDITOR:** Maxine Clarke **EXECUTIVE EDITOR:** Véronique Kiermer
**CHIEF MAGAZINE EDITOR:** Tim Appenzeller
**THIS WEEK/NEWS/FEATURES:** Alison Abbott, David Adam, Ananyo Bhattacharya, Geoff Brumfiel, Declan Butler, Ewen Callaway, Erika Check Hayden, Daniel Cressey, David Cyranoski, Natasha Gilbert, Eric Hand, Heidi Ledford, Corie Lok, Brendan Maher, Emma Marris, Richard Monastersky, Brian Owens, Helen Pearson, Mark Peplow, Eugenie Samuel Reich, Ivan Semeniuk, Quirin Schiermeier, Jeff Tollefson, Richard Van Noorden, Meredith Wadman, M Mitchell Waldrop **COMMENT:** Sara Abdulla, Joanne Baker, Rosalind Cotter, Barbara Kiser, Alison McCook, Lucy Odling-Smee, Sarah Tomlin **NATURE PODCAST/VIDEO:** Adam Rutherford, Thea Cunningham, Geoff Marsh, Kerri Smith, Charlotte Stoddart **NEWS AND VIEWS:** Sadaf Shadan, Ana Lopes, Andrew Mitchinson, Cesar Sanchez, Marian Turner **RESEARCH:** Ritu Dhand, Karl Ziemelis, Francesca Cesari, I-han Chou, Tanguy Chouard, Rosamund Daw, Alex Eccleston, Angela Eggleston, Joshua Finkelstein, Henry Gee, Patrick Goymer, Noah Gray, Marie-Thérèse Heemels, Magdalena Helmer, Karen Howell, Claudia Lupp, Barbara Marte, Andrew Mitchinson, Juliane Mössinger, Deepa Nath, Wayne Peng, Leslie Sage, Magdalena Skipper, Clare Thomas, Maria Trajkovska, John VanDecar, Liesbeth Venema, Ursula Weiss, Michael White **INSIGHTS/REVIEWS/PERSPECTIVES:** Ursula Weiss **CAREERS:** Gene Russo, Karen Kaplan **TECHNOLOGY FEATURES:** Monya Baker **FUTURES:** Colin Sullivan **SUPPLEMENTS:** Herb Brody, Michelle Grayson, Tony Scully **SUBEDITORS:** Colin Sullivan, Mary Abraham, Sarah Archibald, Nicola Bailey, Anne Blewett, Nicolas Fanget, Isobel Flanagan, Paul Fletcher, Sasha Lewis, Dinah Loon, Louisa Lyon, Anna Novitzky, Yukie Ozawa, Katherine Perry, David Price, Jenny Rooke, David Winters, Anna York **PRODUCTION:** Jenny Henderson, Leonora Dawson-Bowling, Dan Foley, Stewart Fraser, Susan Gray, Simon Gribbin, Hazel Mayhew, James McQuat, Emilia Orviss, Yvonne Strong, Robert Sullivan, Roberta Tarran, Charles Wenz **ART AND DESIGN:** Kelly Buckheit Krause, Wesley Fernandes, Madeline Hutchinson, Barbara Izdebska, Paul Jackman, Jasiek Krzysztofiak, Annthea Lewis, Chris Ryan, Nik Spencer, Claire Welsh, Mark Wood **JAPAN EDITION:** Eiji Matsuda, Mika Ishida, Mitsuyo Utsugi, Sayaka Ayame **PRESS OFFICE:** Ruth Francis, Neda Afsarmanesh, Lisa Boucher, Rachel Twinn, Rebecca Walton **ADMINISTRATION:** Paul Sng, Liz Bankes, Roseann Campbell, Peter Cary, Ffion Cleverley, Dana Coles, Faye Fornasier, Ashlee Gagui, Peter Hewitt, Amy Invernizzi, Caroline McLean, Rina Nozawa, Anastasia Panoutsou, Kenneth Simpson, Diane Yorke

## PUBLISHING

**LONDON** **feedback@nature.com**

**MANAGING DIRECTOR & PUBLISHER:** Steven Inchcoombe
**ASSOCIATE PUBLISHER:** Arash Hejazi
**PUBLISHING PROJECT MANAGERS:** Claudia Deasy, Christian Manco

**TOKYO** **feedback@natureasia.com**

**REGIONAL MANAGING DIRECTOR — ASIA-PACIFIC:** David Swinbanks
**DIRECTOR — ASIA-PACIFIC:** Antoine E Bocquet
**OPERATIONS DIRECTOR — ASIA-PACIFIC:** Hiroshi Minemura

## DISPLAY ADVERTISING, SPONSORSHIP & NATUREJOBS

**GLOBAL HEAD:** Andrew Douglas a.douglas@nature.com

### DISPLAY ADVERTISING AND SPONSORSHIP

**GLOBAL HEAD:** Gerard Preston g.preston@nature.com
**ASSISTANT SPONSORSHIP MANAGER:** Reya Silao Tel: +44 (0)20 7843 4977 (r.silao@nature.com)

**EUROPE/REST OF WORLD display@nature.com Tel: +44 (0)20 7843 4960**

**GERMANY/SWITZERLAND/AUSTRIA:** Sabine Hugi-Fürst Tel: +41 52761 3386
**UK/IRELAND/SCANDINAVIA/SPAIN/PORTUGAL:** Evelina Rubio Hakansson Tel: +44 (0)20 7843 4079
**FRANCE/BELGIUM/THE NETHERLANDS/LUXEMBOURG/ITALY/OTHER EUROPE:** Nicola Wright Tel: +44 (0)20 7843 4959

**NORTH AMERICA** **display@natureny.com**

**NEW ENGLAND/ EASTERN CANADA:** Sheila Reardon Tel: +1 617 494 4900
**NEW YORK/MID-ATLANTIC/SOUTHEAST:** Jim Breault Tel: +1 212 726 9334
**MIDWEST:** Mike Rossi Tel: +1 212 726 9255
**WEST COAST/ WESTERN CANADA:** George Lui Tel: +1 415 781 3804
**WEST COAST NORTH:** Kayla McCutchan Tel: +1 415 403 9038

**ASIA-PACIFIC** **display@natureasia.com**

**ASIA-PACIFIC HEAD:** Kate Yoneyama
**JAPAN:** Ken Mikami Tel: +81 3 3267 8765
**GREATER CHINA/SINGAPORE:** Gloria To Tel: +852 2811 7191
**INDIA:** Vikas Chawla Tel: +91 124 288 1057

### SPONSORSHIP

**BUSINESS DEVELOPMENT MANAGER:** David Bagshaw Tel: +1 212 726 9215
**BUSINESS DEVELOPMENT MANAGER:** Patrick Murphy Tel: +1 617 475 9216

### NATUREJOBS

**naturejobs@nature.com**

**EUROPEAN HEAD OFFICE, LONDON:** Tel: +44 (0)20 7843 4961
**EUROPEAN SALES MANAGER:** Nils Moeller (n.moeller@nature.com) **BUSINESS DEVELOPMENT MANAGER, MIDDLE EAST:** Jon Giuliani (j.giuliani@nature.com) **ADVERTISING PRODUCTION MANAGER:** Stephen Russell **NATUREJOBS ONLINE PRODUCTION:** Dennis Chu
**US HEAD OFFICE, NEW YORK:** naturejobs@natureny.com **US SALES MANAGER:** Ken Finnegan
**JAPAN HEAD OFFICE, TOKYO:** Tel: +81 3 3267 8765 **ASIA-PACIFIC HEAD:** Kate Yoneyama
**ASIA-PACIFIC SALES & BUSINESS DEVELOPMENT MANAGER:** Yuki Fujiwara **BUSINESS DEVELOPMENT MANAGER, GREATER CHINA, SINGAPORE:** Gloria To Tel: +852 2811 7191
**INDIA:** Vikas Chawla Tel: +91 1242881057 (v.chawla@nature.com)

## MARKETING & SUBSCRIPTIONS

**USA/CANADA/LATIN AMERICA** **subscriptions@us.nature.com**

Tel: (USA/Canada)+1 866 363 7860; (outside USA/Canada) +1 212 726 9223
**MARKETING:** Sara Girard **FULFILMENT:** Karen Dawson

**JAPAN/CHINA/KOREA** **subscriptions@natureasia.com**

Tel: +81 3 3267 8751 **MARKETING:** Sachiyo Ikeda **FULFILMENT:** Masayo Tanaka

**EUROPE/REST OF WORLD** **subscriptions@nature.com**

Tel: +44 (0)1256 329242
**MARKETING:** Elena Woodstock, Marcus Parker, Hannah Phipps **FULFILMENT:** Jean Burrows

**INDIA** **npgindia@nature.com**

Tel: +91 124 2881053/54
**HEAD OF BUSINESS DEVELOPMENT, INDIA:** Debashish Brahmachari **MARKETING:** Harpal Singh Gill

**ANNUAL SUBSCRIPTIONS (UK/Europe, including post and packing)**

**FULL RATE:** £1,400
**PERSONAL RATES:** UK £135; Europe €209; Rest of World £210
**STUDENT RATES:** UK £79; Europe €122; Rest of World £125
Individual rates apply to payments by personal cheque or credit card. Back issues: £20.00.

**ANNUAL SUBSCRIPTIONS (USA, including post and packing)**

**INSTITUTIONAL/CORPORATE RATE:** $2,730 ; **PERSONAL RATE:** $199; **STUDENT RATE:** $99; **POSTDOC RATE:** $119
Printed in USA. Individual rates available only to subscribers paying by personal check or credit card. Orders for student/postdoc subscriptions must be accompanied by a copy of student ID. Rates apply to USA, Canada, Mexico/Central & South America. Add 7% GST tax in Canada (Canadian GST number 140911595). **BACK ISSUES:** US$20.00.

**ANNUAL SUBSCRIPTIONS (Asia, including post and packing)**

**JAPAN:** **年間購読料金**: *Natureのみ* 437,800 円 +21,890 円(消費税)=459,690 円(合計)
*Nature + Nature Digest* 466,550 円 +23,328 円(消費税)=489,878 (合計)
お問い合わせ:日本出版貿易株式会社 国内営業 1 部
〒101-0064 東京都千代田区猿楽町1-2-1 Tel: 03-3292-3767 Fax: 03-3292-8766
**個人/学生購読について**
NPG ネイチャー アジア・パシフィック フルフィルメント & カスタマーサービス
〒162-0843 東京都新宿区市谷田町2-37 千代田ビル Tel: 03-3267-8759 Fax: 03-3267-8752
**CHINA:** www.natureasia.com/ch/subscribe; **KOREA:** www.natureasia.com/korea/subscribe

## NATURE ONLINE

**WWW.NATURE.COM/NATURE**

**CHIEF TECHNOLOGY OFFICER:** Howard Ratner **ASSOCIATE DIRECTOR, NATURE.COM:** Daniel Pollock **PLATFORM DEVELOPMENT:** Jeremy Macdonald **WEB PRODUCTION:** Alexander Thurrell, Deborah Anthony **WEB DESIGN:** Juan-Carlos Sobrino, Luke Stavenhagen **PLATFORM INTERFACES:** Amanda Ward

## SITE LICENCES, FULFILMENT & CUSTOMER SERVICES

**SITE LICENCES:** npg.nature.com/libraries feedback@nature.com
**CHIEF OPERATIONS OFFICER:** John Carroll **ASSOCIATE DIRECTOR OF OPERATIONS:** Dominic Pettit
**CUSTOMER SERVICE:** Liz Davidson **ORDER PROCESSING OPERATIONS:** Steve Poulter

## NATURE PUBLISHING GROUP OFFICES

**LONDON:** The Macmillan Building, 4 Crinan Street, London N1 9XW Tel: +44 (0)20 7833 4000 Fax: +44 (0)20 7843 4596/7
**BASINGSTOKE:** Nature Publishing Group, Subscriptions, Brunel Road, Basingstoke, Hants RG21 6XS, UK Tel: +44 (0)1256 329242
**WASHINGTON DC:** 968 National Press Building, 529 14th St NW, Washington DC 20045-1938 Tel: +1 202 737 2355
**NEW YORK:** 75 Varick St, 9th Floor, New York, NY 10013-1917 Tel: +1 212 726 9200
**BOSTON:** 25 First Street, Suite 104, Cambridge, MA 02141 Tel: +1 617 475 9275
**SAN FRANCISCO:** 225 Bush Street, Suite 1453, San Francisco, CA 94104 Tel: +1 415 403 9027
**MUNICH:** Nymphenburger Strasse 14, D-80335 München Tel: +49 89 549057-13
**PARIS:** Moulin de Malicorne, Sainte Maure de Touraine, Paris 37800 Tel: +33 2 47 35 72 15
**TOKYO:** Chiyoda Building, 2-37 Ichigayatamachi, Shinjuku-ku, Tokyo 162-0843 Tel: +81 3 3267 8751
**INDIA:** Nature Publishing Group, 3A, 4th Floor, DLF Corporate Park, Gurgaon 122002 Tel: +91 124 2881053/54

---

*Nature*® (ISSN 0028-0836) is published by Nature Publishing Group, a division of Macmillan Publishers Ltd (The Macmillan Building, 4 Crinan Street, London N19XW). Registered as a newspaper at the British Post Office. North and South American orders to subscriptions@natureny.com (Tel: +1 866 363 7860). Other orders to subscriptions@nature.com (Tel. +44 (0)1256 329242; Fax +44 (0)1256 812358). Authorization to photocopy material for internal or personal use, or internal or personal use of specific clients, is granted by *Nature* to libraries and others registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided the relevant copyright fee is paid direct to CCC, 222 Rosewood Drive, Danvers MA 01923, USA. Identification code for *Nature*: 0028-0836/03. CPC PUB AGREEMENT #40032744. In the US, *Nature* (ISSN 0028-0836) is published weekly on Thursday, except the last week in December by Nature Publishing Group, 75 Varick St, 9th Fl, New York NY 10013-1917, USA. US Periodicals postage paid at New York NY, and additional mailing post offices. US POSTMASTER: send address changes to *Nature*, Subscription Dept, 75 Varick St, 9th Fl, New York NY 10013-1917, USA. Published in Japan by NPG Nature Asia-Pacific, Chiyoda Building, 2-37 Ichigayatamachi, Shinjuku-ku, Tokyo 162-0843, Japan. © 2012 Macmillan Publishers Limited. All rights reserved.

# Handcuffed to your pipette?

## Free yourself!

**The Apollo 324™ system automates library preparation for all major NGS systems, so you can spend your time on more important things. Setting up and starting a run takes only 15 minutes of hands-on time.**

**Then you are free.**

Our PrepX™ reagent kits generate consistent, quality libraries for Ion Torrent PGM™, Roche 454™ GS FLX Titanium™, FLX+ and GS Junior™, and all Illumina® platforms. Color-coded reagent kits — for DNA, RNA and ChIP-seq applications — minimize errors and simplify instrument setup. Pre-aliquoted strips let you set up full or partial runs with no reagent waste.

integenX

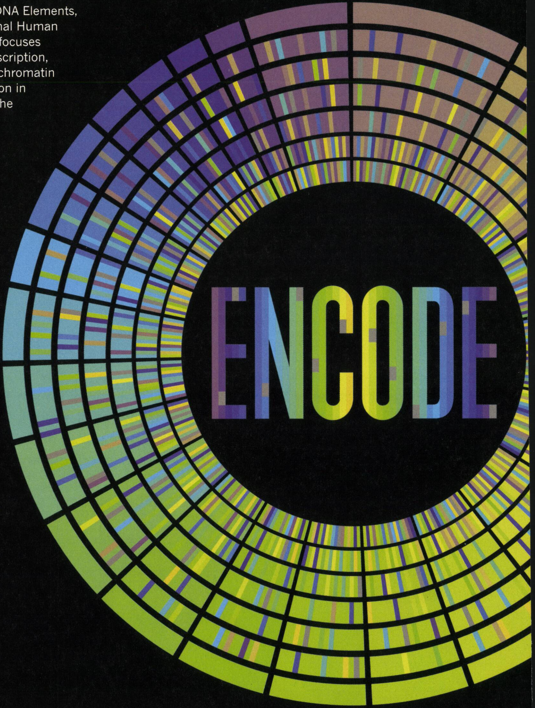
**For more info visit: www.integenx.com • 888 371 4749**

Copyright © 2012 IntegenX Inc. All rights reserved. Patents pending. FOR RESEARCH USE ONLY.
Product specifications may change. IntegenX, Apollo 324 and PrepX are trademarks of IntegenX Inc.
All other brands and names are the sole property of their respective owners.

nature | ENCODE

ENCODE, the Encyclopaedia of DNA Elements, is a project funded by the National Human Genome Research Institute and focuses on identifying all regions of transcription, transcription factor association, chromatin structure and histone modification in the human genome sequence. The main findings of ENCODE are now presented in 30 papers published simultaneously in *Nature*, *Genome Research* and *Genome Biology*.

Access videos, Features and the collected research papers, and explore the thematic threads that run through them via the *Nature* ENCODE explorer or the *Nature* ENCODE app.



Produced with support from

illumina®

Available on the App Store

nature.com/ENCODE

nature publishing group npg

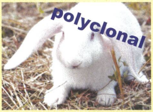
**Davids Biotechnologie**

Davids GmbH Biotechnologie

**Antibodies made in Germany**

**Your manufacturer since 1996**

monoclonal



**www.davids-bio.de**

naturejobs

**Attention Employers!**

Stay up to date with deadlines and upcoming ***Naturejobs*** features:

Sign up for **FREE** employer e-alerts: **www.naturejobs.com/ealerts**

SIRION BIOTECH

**INDUCIBLE GENE KNOCKDOWNS OF NEARLY 100%**

With a Satisfaction Guarantee? Is that possible?

**RNAiONE™**
Europe's proven shRNA validation platform

secure an extra **15%** off until October 31st, 2012

www.sirion-biotech.com

RNAiONE™

nature MIDDLE EAST
Emerging science in the Arab world

twitter facebook google+

RESEARCH, NEWS, JOBS, BLOGS

The Arab world has a rich history of scientific breakthroughs. Ensure you're part of the scientific and medical community gaining access to **the latest research in the region**.

From the latest news, in-depth features, expert commentaries and research highlights to regional jobs, event listings and alerts direct to your inbox – ***Nature Middle East*** is your free comprehensive portal keeping you at the centre of science in the Arabic speaking Middle East.

**nature.com/nmiddleeast**

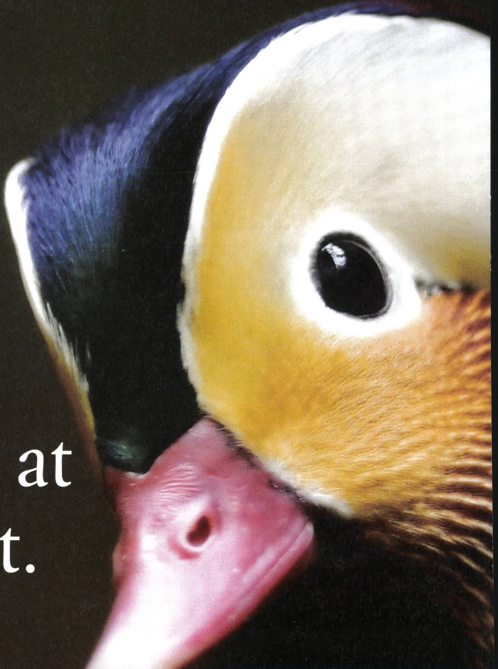
Download the free app for your phone at **http://gettag.mobi**

Sponsored by

nature publishing group npg

NEW ENGLAND
BioLabs® Inc.
*enabling technologies in the life sciences*

For more information, including our FREE SAMPLE special offer, visit
**www.Q5PCR.com**



Fidelity at its finest.

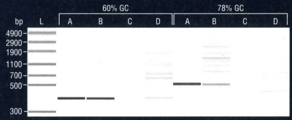
Q5™ High-Fidelity DNA Polymerase

Q5 High-Fidelity DNA Polymerase sets a new standard for both fidelity and performance. With the highest fidelity amplification available (>50X higher than *Taq*), Q5 DNA Polymerase results in ultra-low error rates. Its unique buffer system provides superior performance for a broad range of amplicons, regardless of GC content. Available in master mix and hot start formulations, Q5 DNA Polymerase represents the finest in fidelity.

ALSO AVAILABLE: Optimized NEBNext® formulation for next generation sequencing library amplification

Mandarin Ducks *(Aix galericulata)* are frequently featured in Chinese art and are regarded as a symbol of fidelity.

Robust amplification even with high GC amplicons



*Amplification of two human genomic amplicons of mid to high GC content. All reactions were conducted using 30 cycles of amplification and visualized by microfluidic LabChip® analysis. All polymerases were cycled according to manufacturer's recommendations. For the 78% GC amplicon, GC Buffers or enhancers were used when supplied with the polymerase.*

*A = Q5™ High-Fidelity DNA Polymerase (NEB)*
*B = Phusion® High-Fidelity DNA Polymerase (NEB)*
*C = KOD DNA Polymerase (EMD)*
*D = PfuUltra™ High-Fidelity DNA Polymerase (Agilent)*

PHUSION® is a registered trademark and property of Thermo Fisher Scientific.
Phusion® DNA Polymerase was developed by Finnzymes Oy, now a part of Thermo Fisher Scientific.
PFUULTRA™ is a trademark of Agilent Technologies, Inc.
LABCHIP® is a registered trademark of Caliper Life Sciences, part of Perkin Elmer, Inc.
Q5™ is a trademark of New England Biolabs, Inc.

# THIS WEEK

## EDITORIALS

**WORLD VIEW** How to mend the broken pieces of research **p.7**

**ATTACHED** On the sticky secrets of barnacle glue **p.8**

**ENERGY** Hunter-gatherer study questions obesity **p.9**

# Accountable and transparent

*The US government has changed how biomedical scientists disclose their financial interests. The revised rules are welcome, but Internet access to the identified conflicts should be a requirement.*

Toughened rules for how US biomedical scientists report financial interests came into force last month. The changes, which affect scientists who receive grants from the government, are welcome — although in one respect they do not go far enough.

About 38,000 researchers, most of them recipients of grants from the US National Institutes of Health (NIH), the world's largest medical-research funder, will need to comply with the beefed-up rules. The changes update regulations put in place in 1995 to ensure that investigator bias doesn't sway the design, conduct or reporting of research.

There are several important changes. First, investigators must now disclose to their institutions every "significant financial interest" belonging to themselves or their immediate family that is related to any of their institutional responsibilities — from teaching and seeing patients to lab research and service on ethics committees. This requirement appropriately casts a broader net than the previous rules, which generally asked for disclosure on only a project-specific basis.

The change ends ambiguity that, for instance, might have allowed a researcher to conclude that paid service on the board of a major pharmaceutical company drew only on clinical expertise, and therefore was not relevant to a government-funded research project that used one of the company's experimental compounds. Under the updated rules, there will be no question that such income must be disclosed, and institutions will have a more complete picture of their scientists' potentially relevant financial interests.

It takes only one example to drive home the significance of this change. Between January 2000 and January 2006, high-profile psychiatrist Charles Nemeroff, then at Emory University in Atlanta, Georgia, received more than US$800,000 in payments from drugmaker GlaxoSmithKline for over 250 speeches that he gave to psychiatrists. He failed to disclose this income to Emory administrators. After being discovered, Nemeroff argued that the rules on whether such income was reportable were ambiguous.

The tougher rules, crucially, give institutions prime responsibility for determining whether a given financial interest — company-paid speaking honoraria, consulting fees, paid authorship, travel reimbursements and stock ownership all qualify — is related to a government-funded grant, and whether it constitutes a conflict. Under the old regime, the scientist was charged with deciding whether a given interest was related to the research and thus whether it was reportable. That arrangement did not inspire confidence — a problem in an era in which public trust in the medical enterprise is at risk and must be built, not undermined.

The updated rules also lower the threshold at which an interest is defined as significant, from $10,000 under the old rules to $5,000. In a moribund economy with many US taxpayers struggling to make ends meet, this is fitting.

The rules have also been strengthened in other important ways. For instance, far more detail will now be reported by institutions to the NIH about each identified conflict, including the approximate dollar value of the interest and the measures being taken to manage or eliminate the conflict. There is also, importantly, an explicit exception to the disclosure requirements for income that scientists earn from universities or government agencies for teaching, serving on advisory or review panels and giving seminars or lectures.

The new rules fall down, however, in one significant regard. When it first published the proposed changes, the NIH described what it called "an important and significant new requirement to … underscore our commitment to fostering transparency, accountability, and public trust".

That requirement was that institutions would post details of their investigators' financial conflicts of interest on a publicly accessible website that was updated every year. In the final iteration of the new rules, the website has been made optional, and institutions faced with requests for information may instead respond in writing, within five business days. This is an outdated approach to transparency. It will not advance the public's faith in timely, comprehensive and truly accessible disclosure, at a time when the boundary between academia and industry has become ever more porous, and when the average citizen's trust in government-funded medical research is ever more crucial. The NIH should revise the rules again to make the website mandatory. It is within the agency's power to insist on this standard, and it is the right thing to do. ■

***"Public trust in the medical enterprise is at risk and must be built, not undermined."***

# Spinning threads

*Publication of ENCODE data drives innovation in data mining.*

There can be few scientists who have not used a brightly coloured highlighter pen to mark the most interesting parts of a research paper, report, proposal or (librarians look away) book. It is a natural reaction when faced with a swamp of information — to build islands of focus that can be identified and linked, both in print and in the mind.

This week, *Nature* introduces a new concept in the publishing and dissemination of scientific information: one that is a response to the increasing complexity of modern research, and one that draws heavily on the contribution of the humble highlighter.

Starting on page 45, we publish a package of material that centres on the results from the ENCODE consortium, including 6 of the 30 papers the project has produced. The ENCODE — Encyclopedia of DNA Elements — consortium set out to describe all the functional elements in the human genome. Their headline conclusion: more than 80% of the human genome's components have now been assigned at least one biochemical function.

The six papers that *Nature* publishes (the others appear simultaneously in *Genome Research* and *Genome Biology*) may look like conventional research reports, but in the digital world they begin to take on new form — as themed threads. If you are reading this online, then click on the link. If you are reading it in print, then have a look at the version on *Nature*'s ENCODE explorer website (www.nature.com/encode) or, better still, the iPad app.

As part of the publication process, the ENCODE authors asked for something extra: to select and package together the sections from each paper that will be of particular interest to scientists in various and varied fields. Just as a postdoctoral researcher looking at transcription factors would use a highlighter to mark up different bits of the papers from, say, a colleague looking at DNA methylation, so the ENCODE authors thought that researchers across the biological spectrum would want to be able to pull together pieces from each of the digital versions that were of specific interest to them. Our editors agreed, and the result is 13 online threads — biological themes that contain no original material but instead harvest and combine related paragraphs, figures and tables from the 30 papers.

The threads, we hope, will help readers to make sense of the dizzying amounts of data produced during the five years of the main ENCODE effort. And they should allow scientists to exploit more easily the information in their own studies, and that, after all, was the point of the project in the first place. Presented online, the threads are filled with links that allow readers to jump easily from paper to paper, to see where the information comes from and how the data are interconnected.

Alongside the thread concept, the ENCODE package introduces another technical innovation, at least one new to *Nature*. Using a 'virtual machine', online readers can access software designed to perform set computational functions on some of the ENCODE data.

The idea is to allow readers to recreate the analyses behind the specific aspects of the paper and to see how the outcome changes when specific parameters are tweaked. Think of it as a bridge that links the data, the analysis and the relevant description and discussion in the formal papers.

We are eager to hear what readers and users of the material think of these approaches. If they are useful, and early feedback suggests that they will be, then scientists who work on other similarly data-rich and analysis-heavy projects should take note. Results from projects that aim to sequence the human microbiome or different forms of cancer, for example, produce piles of data that could be split along many different themes, and so divided into threads. In many cases the true hard work — the science — is done. Threads, then, are just a different way to package the results.

> ***"Scientists who work on other data-rich and analysis-heavy projects should take note."***

Some practical problems remain in applying these ideas more widely. The thread concept depends on cooperation between publishers, as well as open access to the papers and appropriate copyright agreements. And the virtual machine demands well curated data that are available to all.

Why are there 13 ENCODE threads? Good question, there could have been many more — as many as there are questions raised in the minds of scientists by the mass of information that the project has placed at their disposal. If your particular interest or angle is not already selected and presented as a theme, then apologies — there is always the old-fashioned route: print the papers and attack them with a highlighter. ■

---

# Moonlight drive

## *The data from the ageing Voyager probes are illuminating the edge of the Solar System.*

Someone in the NASA media-relations office knows their music. A press release from the agency last month stated that the twin Voyager spacecraft were poised to Break on Through to the Other Side — referring to the probes' approach to the edge of the Solar System, but also to a 1967 hit from the US band The Doors. NASA pointed out to journalists that the missions were launched 35 years ago and was no doubt hoping for some (more) positive coverage to mark the anniversary. What's more, on 13 August, Voyager 2 became the longest-operating spacecraft, beating the record of Pioneer 6, which was launched in December 1965 and returned its final signal some 12,758 days later. (Voyager 2, counterintuitively, was launched two weeks before Voyager 1, but the latter is now the farthest from the Sun.)

The spin doctors can be excused this time. Voyager is a truly great mission, and one that reporters still find hard to resist — some of them have been happily writing about its discoveries ever since the two craft launched in 1977. It is the science story that keeps on giving: the deep, hazy atmosphere of Saturn's moon Titan; the volcanoes of Jupiter's moon Io; the large, unusual magnetic field of Uranus; and the geysers of Triton, the frozen world that orbits Neptune — all discovered and lapped up by an eager public as the probes skimmed past the outer planets.

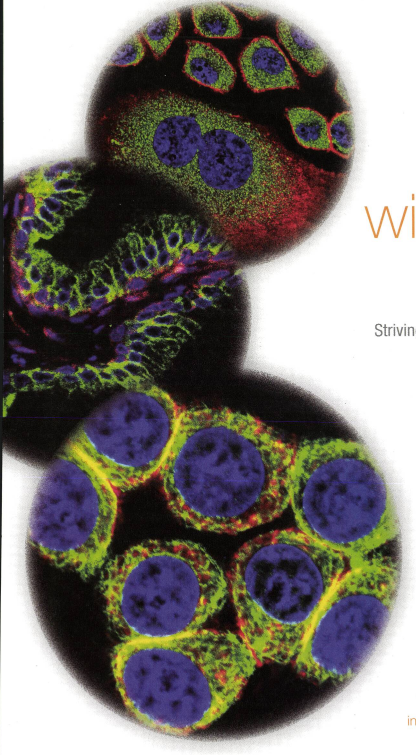
Still, their work is not done. Even though the probes are now more than 15 billion kilometres away from the Sun, their handlers on the ground remain in near-daily contact, as the spacecraft continue to send back useful information — now about the farthest reaches of the Solar System. Last year, NASA even coaxed the ageing and radiation-blasted parts of Voyager 1 into performing a series of rolls to have a proper look around. It was curious because some of the data being sent back from the spacecraft seemed to suggest that the edge of the Solar System was nearby. Levels of high-energy cosmic rays, which originate far beyond our corner of space, had spiked. And the number of lower-energy particles that come from closer to home seemed to dip.

The results of the latest tests, which are published on page 124, have surprised many. If Voyager 1 truly is near the point where the heliosphere — the bubble of charged particles from the Sun — fades to interstellar grey, then it should have found solar particles that have been buffeted by the winds of deep space, generated by supernovae that exploded long ago elsewhere in the Galaxy. In fact, the particles it found had effectively been becalmed.

The implications of the discovery for our understanding of the structure of the Solar System, and how it changes as it whizzes through space, are profound. As a News story on page 20 explains, the find could mean that astronomers will have to rethink their models of the heliopause, the boundary at which the outward pressure of the heliosphere is balanced by the inward push of outer space. Or it could mean that Voyager 1 is still some distance from the heliopause.

That would no doubt disappoint the NASA press office, which is eager to announce that at least one probe has entered a new realm of discovery — and before the batteries of the spacecraft run out, in a decade or so. But it should not lose heart. Like the Voyager probes, The Doors are still going, albeit not as strongly and with their best work probably behind them. If the heliopause is farther away than we thought, and the reach of the solar wind longer than we realized, then the Voyager twins still have many years remaining as Riders on the Storm, and some way to go before they reach The End. ■

➲ **NATURE.COM**
To comment online, click on Editorials at:
go.nature.com/xhunqv



# Proteome-wide antibody coverage.

**Striving towards an antibody for each human protein.**

Unsurpassed development capabilities built over the last ten years allow Abgent to produce more than 5,000 new antibodies per year. Protein A and peptide affinity purification, along with rigorous antibody validation, ensures antibody quality and assay flexibility. The end result is a broad offering of in-house developed, high quality antibodies tailored to your research and drug discovery needs.

**Applications:** Western Blot, Immunohistochemistry, Immunofluorescence, Flow Cytometry, ELISA, and more.

**20µg Evaluation Size Vials**
available for most new products.

Visit our website at **www.abgent.com** for more information on Abgent's product and service offerings.

© 2012 ABGENT. All rights reserved. It's all about New. is a trademark of ABGENT. For Research Use Only. Not For Use In Diagnostic Procedures. ABG-00104-0612

**It's all about New™**
**Unsurpassed Development Capabilities**

ABGENT ANTIBODIES ARE ESSENTIAL TOOLS FOR THE STUDY OF CANCER, NEURODEGENERATIVE DISEASES, METABOLIC SYNDROME, IMMUNOLOGY, APOPTOSIS, AUTOPHAGY, DEVELOPMENT AND MORE.

**Visit our website for a complete product listing.**

www.abgent.com

药明康德
WuXi AppTec

ABGENT

introducing:

LEXOGEN
Enabling complete transcriptome profiling

SENSE™

Stranded libraries for RNA sequencing

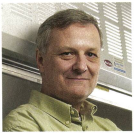
Lexogen's new
SENSE mRNA-Seq
library prep kit

- Easy to use
- Generates ready to sequence libraries from total RNA in under 4 h
- Eliminates rRNA contamination
- **Maintains >99.9% strandedness**
- Empowers your transcriptome research and ability to discover.

For more information visit us at www.lexogen.com.

# WORLD VIEW

*A personal take on events*


MOUNT SINAI HOSPITAL

# We must be open about our mistakes

*Greater transparency about the scientific process and a closer focus on correcting defective data are the way forward, says* **Jim Woodgett**.

There is increasing unrest in global science. The number of retractions is rising, new examples of poor oversight or practice are being uncovered and anxiety is building among researchers. Those of us who work in the life sciences are discovering that some of our basic premises are flawed or inaccurate — cell lines have been misidentified and drug metabolism in animal models misjudged. Even high-profile findings have been questioned. Building on solid foundations was an architectural principle understood by the ancient Greeks and Egyptians, yet we may be constructing our castles on swampland. Is it a surprise that clinical translation fails so often?

Although most mistakes are unintentional and sometimes unavoidable, there are also deliberate efforts to deceive. Scientists (especially those of us in biomedical research) must do more to detect and be seen to correct errors as an on-going imperative.

We scientists must recognize that, to the public and politicians, we are a privileged and elite group. The products of our work are largely incomprehensible to non-experts — and even to colleagues on the periphery of the same field. Like an iconoclastic gentlemen's club, our community has developed rules and etiquette to maintain order. But, unlike a club, our sponsorship fees are paid by taxpayers and philanthropic donations.

The scientific community must be diligent in highlighting abuses, develop greater transparency and accessibility for its work, police research more effectively and exemplify laudable behaviour. This includes encouraging more open debate about misconduct and malpractice, exposing our dirty laundry and welcoming external examination. A good example of this, the website Retraction Watch (retractionwatch.wordpress.com), shines light on problems with papers and, by doing so, educates and celebrates research ethics and good practice. Peer pressure is a powerful tool — but only if peers are aware of infractions and bad practice.

SEARCHABLE **INTERNET** TECHNOLOGIES, SUCH AS SOCIAL MEDIA AND BLOGS, ARE HELPING TO LIFT THE VEIL OF SECRECY **OVER SCIENCE.**

We might also better foster and acknowledge aspects of research that are often overlooked. Efficient reagent exchange and sharing, for example, protects against cheats and can help to correct more common, unintentional errors.

The inherent uncertainty of research provides a safe haven for data omission, manipulation or exaggeration. Because interpretation of data is an imperfect science, there are few consequences for those tempted to oversell their findings. On the contrary, such faulty embellishment can help to determine whether a study is published — and where. Moreover, because failure to reproduce a published finding can be due to innocent factors, significant errors or falsehoods may be overlooked or simply pass unchallenged. As a result, modern science can churn out a flotsam of dead-end data that pollute the literature and waste precious resources.

➲ **NATURE.COM**
Discuss this article online at:
go.nature.com/ookutx

To counter this, barriers to correction of the public record should be low but rigorous. Publication of refutations or modifications should be encouraged by journals and funding agencies. One may argue that if a study is ignored it does no harm, but superfluous publication clutter is not benign. Minimally, it adds chaff to the wheat, but it also promotes mediocrity by example. More importantly, it provides meticulously documented evidence of apparent waste to funders and the public.

In a culture of publish or perish, the continuing growth in the number of scientific journals is hardly a surprise. But does this proliferation of papers reflect better science, or merely dilution? When a third of all papers are never cited, it is reasonable to question why so many are published. If the answer is simply as a form of accepted currency to indicate productivity, then our evaluative systems must become less reliant on publication quanta.

Before we complain legitimately about grant success rates and funding pressures, we must ensure that our own house is in order. The act of publishing takes significant effort, yet we still publish low-impact studies as the required unit of research. We must learn to stop publishing everything and find other ways to document and recognize our studies, such as searchable publication of theses, meeting proceedings and posters.

And take the way most scientists access money from the public purse. Despite being the conduit to research funds, grant proposals undergo limited vetting of their content. Unlike manuscripts that pass peer review, these documents are treated as confidential, so their writers are difficult to hold to account. There are legitimate concerns about intellectual property and fear of being scooped by competitors, but why not make such documents public after a period of time? Indeed, some scientists are already publishing their grant applications on the Internet, ostensibly to help educate new researchers. But this also allows validation and cross-checking and sets a new bar for transparency.

Other searchable Internet technologies, such as social media, blogs, slide-sharing sites and even video-sharing sites such as YouTube, are helping to lift the veil of secrecy over science. This increased transparency, associated with wider access and discussion, is a powerful weapon to reduce scientific misinformation of all sorts — and one that all honest and careful scientists should embrace. Transgressions and errors will be more quickly detected and more widely communicated when more of what we do is exposed to scrutiny. As security professionals know, the surveillance camera does not need to be turned on to deter. ■

**Jim Woodgett** *studies signalling pathways at the Samuel Lunenfeld Research Institute in Toronto, Canada.*
*e-mail: woodgett@lunenfeld.ca*

# RESEARCH HIGHLIGHTS

*Selections from the scientific literature*

CHEMISTRY

## When alkanes turn tail

Alkanes are molecules that contain only carbon and hydrogen atoms, connected by single bonds. Short-chain alkanes such as butane and octane — which contain linear chains of four and eight carbon atoms, respectively — stretch out in extended zig-zags. However, longer hydrocarbon chains tend to fold themselves into hairpin structures.

Ricardo Mata, Martin Suhm and their colleagues at the University of Göttingen, Germany, determined the point at which this transition becomes energetically favourable. The researchers performed spectroscopy on supersonic jets of alkane molecules at temperatures of 100–150 kelvin — and found that the folded structure becomes more stable than the extended conformation when an alkane chain is around 18–19 carbon atoms long.

The result broadly agrees with the authors' quantum calculations, and can be used to train computer models of molecular mechanics.

***Angew. Chem. Int. Edn*** **http://dx.doi.org/10.1002/anie.201202894 (2012)**

R. HODDINOTT/NATUREPL.COM

MATERIALS

## Why barnacles stick around

Barnacles are among the clingiest of creatures, but how they manage to stick so tenaciously to surfaces is unclear.

When Jaimie-Leigh Jonker of the National University of Ireland, Galway, and her colleagues examined the barnacle *Lepas anatifera*, they found that its adhesion system is radically different from that of other clingy sea creatures, such as mussels and tubeworms. Large, single-cell glands in *L. anatifera* secrete a clumpy substance filled with sticky proteins, although exactly how the glue works remains mysterious.

Researchers hope that future studies of barnacle glue will yield better adhesives, particularly for medical applications.

***J. Morphol.*** **http://dx.doi.org/10.1002/jmor.20067 (2012)**

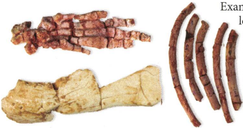
PALAEONTOLOGY

## Excavation of a digger

Examination of a 57-million-year-old nearly complete fossil skeleton (**selected bones pictured**) has advanced a long debate over the place of the mammal *Ernanodon antelios* in evolutionary history.

The fossil of the ancient mammal was discovered in rocks in Mongolia. Peter Kondrashov and Alexandre Agadjanian from the Borissiak Paleontological Institute of the Russian Academy of Sciences in Moscow describe *E. antelios* as having strong forelimbs and large claws, which it used to scratch and dig for food. Examination of the bones led the authors to suggest that the mammal is more closely related to pangolins than it is to armadillos and anteaters.

***J. Vertebr. Paleontol.*** 32, **983–1001 (2012)**



P. KONDRASHOV

OCEAN BIOCHEMISTRY

## The mystery of high seas methane

Marine microbes offer a plausible explanation for the surprising abundance of methane in oxygenated parts of the ocean.

Scientists have previously theorized that ocean methane might be a by-product of microorganisms' use of methylphosphonic acid as a source of phosphorus. But it was unclear where the acid itself came from. William Metcalf and Wilfred van der Donk at the University of Illinois in Urbana and their colleagues show that a microbe called *Nitrosopumilus maritimus* carries genes that encode a pathway for methylphosphonate synthesis.

A crucial gene in this pathway is also found in many other marine microbes, suggesting that these organisms may be the source of the unexplained ocean methane.

***Science*** 337, **1104–1107 (2012)**

EVOLUTIONARY ANTHROPOLOGY

## Small families in rich societies

The tendency of families in wealthier societies to produce fewer children is hard to explain in evolutionary terms. A study of Swedish families examines the paradox, known as demographic transition.

One model proposed to explain the phenomenon holds that fewer offspring receive more resources, making them more likely to have offspring themselves. The model posits that richer people might have fewer children, but would ultimately have more descendents over subsequent generations.

Not so, say Anna Goodman of the London School of Hygiene and Tropical Medicine and her team. In their analysis of 14,000 Swedish people born between 1915 and 1929 and their descendents, small family size predicted greater socioeconomic success in children, grandchildren and great-grandchildren, particularly among families that already had high socioeconomic status. But small family size did not translate into greater reproductive success among the descendants.

***Proc. R. Soc. B*** **http://dx.doi.org/10.1098/rspb.2012.1415 (2012)**

**BOTANY**

## Plants split cells to put down roots

Plants cells cannot migrate, so plants control the development of multilayered tissues such as roots through asymmetric cell divisions that create layers with different identities and functions.

A team headed by Athanasius Marée of the John Innes Centre in Norwich, UK, and Ben Scheres at the University of Utrecht in the Netherlands unravelled the molecular pathway that regulates these cell divisions in the root tip. Stem cells in the model plant *Arabidopsis* are triggered to divide unevenly by a positive feedback loop that takes effect when a protein called RETINOBLASTOMA ceases to inhibit another, called SCARECROW. Gradients of a growth hormone and a protein called SHORT ROOT ensure that this loop is triggered in the correct place. Protein degradation during the division prevents the process from continuing indefinitely.

***Cell*** **http://dx.doi.org/10.1016/j.cell.2012.07.017 (2012)**

**ASTROPHYSICS**

## Disintegrating planet spotted

NASA's Kepler spacecraft seems to have spotted a distant, rocky planet that is falling apart.

Kepler hunts for planets beyond the Solar System by searching for steady, periodic dimming in the light of parent stars, which indicates the passage of an orbiting body. In the case of the star KIC 12557548, however, the drop in starlight varies in strength with each passage. Scientists have suggested that this variability is a sign of an orbiting planet that is trailed by a large dust cloud.

Matteo Brogi of Leiden University in the Netherlands and his team modelled the dust cloud and found that its presence could indeed explain the Kepler data. The cloud is probably the result of the planet being bombarded by so much stellar radiation that it has begun breaking up into dust.

***Astron. Astrophys.*** **http://dx.doi.org/10.1051/0004-6361/201219762 (2012)**

**BIOGEOSCIENCES**

## Pruning back carbon estimates

Incorporating tree-height data into calculations of the amount of carbon stored in tropical forests reduces the estimates by roughly 13%.

Ted Feldpausch of the University of Leeds, UK, and his team analysed data from 327 tropics-wide plots, as well as 20 sites where tropical trees have been cut down, collecting data on factors such as the weight and height of the trees, and their carbon density. The team found that information on tree height was crucial for making accurate biomass estimates, and that the relationship between height and carbon storage varied by region.

The authors underscore the importance of including better data in biomass maps, in which field measurements are increasingly being integrated with remote-sensing data to improve accuracy.

***Biogeosciences*** 9, **3381–3403 (2012)**

## COMMUNITY CHOICE

*The most viewed papers in science*

**ANTHROPOLOGY**

## Hunter-gatherer workout disproved

**✪ HIGHLY READ**
on www.plosone.org in August

Despite their very different lifestyles, a hunter-gatherer expends about the same amount of energy each day as the average person in Europe or the United States.

For 11 days, Herman Pontzer of Hunter College in New York and his colleagues measured daily energy expenditure and physical activity levels in 30 adults from a Hadza hunter-gatherer group in Tanzania. Controlling for factors such as age, sex, body fat and body mass, the researchers compared their results to individual and population data from a spectrum of societies, including Western countries. Hadza individuals walk longer distances and forage for resources. So, unsurprisingly, they had higher physical-activity levels than Westerners. However, on average, both groups used the same amount of energy on a daily basis, as well as when walking or resting, suggesting that the rate of energy expenditure is an evolved trait that is independent of culture.

Obesity trends in Western populations could be unrelated to a sedentary lifestyle, the researchers suggest.

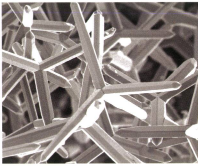
***PLoS ONE*** 7, **e40503 (2012)**

**MATERIALS**

## Sticking the unstickable

Researchers have succeeded in sticking together two supremely unsticky polymers — Teflon and cross-linked poly(dimethylsiloxane), the slippery coating used as backing paper for stickers.

The secret to their success lies in tetrapodal zinc oxide crystals: micrometre-scale structures (**pictured**) shaped rather like children's jacks. Strewing these between the polymers and heating the resulting sandwich to 100 °C for 40 minutes creates a kind of 'micro/nano Velcro'. The polymers can be peeled apart only by applying a force of about 200 Newtons per metre — more than that required to unstick Scotch tape.



ADV. MATER.

Rainer Adelung and his team at the University of Kiel, Germany, did not stick the unstickable for glory alone. Stuck together, these surfaces will have applications in technologies such as membranes for separating liquids, and biomedical implants.

***Adv. Mater.*** **http://dx.doi.org/10.1002/adma.201201780 (2012)**

**➲ NATURE.COM**
For the latest research published by *Nature* visit:
**www.nature.com/latestresearch**

# SEVEN DAYS *The news in brief*

**RESEARCH**

### Genome decoded

The Encyclopedia of DNA Elements (ENCODE) consortium this week publishes the fruits of its endeavour to understand how human cells use the genomic code. Across 30 papers published in *Nature* (see page 45), *Genome Research* and *Genome Biology*, the team reveals that more than 80% of the human genome's components have now been assigned at least one biochemical function. See nature.com/encode for more.

### Resistance warning

More than 40% of multidrug-resistant (MDR) tuberculosis infections are also resistant to some of the common second-line backup drugs, according to research published on 29 August (T. Dalton *et al. Lancet* http://doi.org/h8r; 2012). MDR strains are not routinely screened for resistance to second-line drugs in the poor countries where the incidence of tuberculosis is highest. Out of 1,278 people with MDR tuberculosis, 6.7% could be classified as having extensively drug-resistant tuberculosis — almost untreatable strains that are resistant to several common backup drugs. See go.nature.com/dklimh for more.

### Virus discovery

A new type of phlebovirus causing fever, severe fatigue and nausea has been identified in a paper published on 30 August (L. K. McMullan *et al. N. Engl. J. Med.* **367,** 834–841; 2012). Found in Missouri, it is the first virus pathogenic to humans to be discovered in the United States since hantavirus in 1993. Dubbed the Heartland virus, the phlebovirus is probably spread by the lone star tick (*Amblyomma americanum*) and is distantly related to a tick-borne and potentially lethal phlebovirus discovered in China last year. The two Missouri men infected with the virus recovered, however.

J. FREUND/NATUREPL.COM

## One in five invertebrates face extinction

The first comprehensive effort to review the conservation status of the world's invertebrates shows that about one-fifth of species are at risk of extinction, according to a report from the Zoological Society of London. Such creatures are thought to represent around 99% of the biodiversity on Earth. The report suggests that the greatest threat is to freshwater invertebrates, followed by terrestrial and marine invertebrates, such as nudibranch sea slugs (*Hypselodoris kaname*, pictured). See go.nature.com/r2uf2y for more.

**BUSINESS**

### Drug hope dashed

Prospects for a new class of drug to treat schizophrenia were scotched on 29 August, when pharmaceutical giant Eli Lilly halted the late-phase clinical trial of its drug pomaglumetad methionil, also known as mGlu2/3, which modifies glutamate neurotransmission in the brain. The company, based in Indianapolis, Indiana, said the drug seemed to be ineffective. Current schizophrenia drugs work primarily by reducing levels of the neurotransmitter dopamine in the brain, but they do not control all symptoms of the illness.

### Late apology

Pharmaceutical company Grünenthal, based in Aachen, Germany, has apologized for the first time for the effects of the drug thalidomide. The firm developed the drug, which was used to treat morning sickness in pregnant women between 1957 and 1961. Thalidomide was withdrawn after causing birth defects in thousands of babies.

**FUNDING**

### ArXiv boost

The arXiv preprint server at Cornell University Library in Ithaca, New York, is to get up to US$350,000 a year for the next five years from the Simons Foundation, a charity based in New York that supports basic research. The sum includes an unconditional annual grant of $50,000, with the remainder depending on matching funds from arXiv's other donors, the library said on 28 August. The foundation was set up in 1994 by mathematician and hedge-fund manager James Simons and his wife, Marilyn. See go.nature.com/xfapfr for more.

**POLICY**

### Carbon trade grows

Australia announced on 28 August that it is to join the European Union (EU) Emissions Trading System, marking the first time that a non-European country has linked up with the greenhouse-gas-reduction strategy. Australian firms will be able to cover up to 50% of their carbon emissions by purchasing carbon permits issued to European companies from 2015. EU companies will be able to buy Australian permits from 2018.

## Forest code final

Brazil's controversial forest-protection law reached what is likely to be its final form on 29 August, after a congressional committee made further changes to the version proposed by President Dilma Rousseff in May. The text further reduces protection for forests abutting rivers, for example. See go.nature.com/34qwnl for more.

## Embryo ruling

The European Court of Human Rights ruled on 28 August in favour of an Italian couple who want to be able to screen their *in vitro* fertilized embryos for a disease-causing gene before implantation. A 2004 Italian law currently bans preimplantation genetic diagnosis. The couple both carry mutations that cause cystic fibrosis, and their first daughter has the disease.

## Wolves delisted

The US Fish and Wildlife Service has removed grey wolves from the endangered-species list for Wyoming, the last state in which hunting of the animals was regulated by the federal government (see go.nature.com/4zmmic). Wolves will be managed by the state from 30 September, which will probably mean that wolves can be shot on sight outside protected areas such as Yellowstone National Park. Environmental groups have promised legal action to reverse the move.

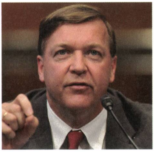
PEOPLE

## Misconduct verdict

Shane Mayack, a former postdoctoral researcher at the Joslin Diabetes Center, an affiliate of Harvard Medical School in Boston, Massachusetts, duplicated figures in two stem-cell papers and poached figures from other sources, an official investigation by the US Office of Research Integrity has concluded. The papers (S. R. Mayack *et al. Nature* **463,** 495–500; 2010, and S. R. Mayack and A. J. Wagers *Blood* **112,** 519–531; 2008), had already been retracted by co-author Amy Wagers, a stem-cell biologist at Joslin and Mayack's mentor. See go.nature.com/jzdtny for more.

## Biosecurity leader

Samuel Stanley, president of Stony Brook University in New York, will serve as chair of the US National Science Advisory Board for Biosecurity. The board has been enmeshed in controversy for recommending in December 2011 that two research papers on highly pathogenic avian influenza H5N1 be redacted for safety and security reasons, before finally voting in favour of full publication in March this year. Stanley (**pictured**) replaces acting chair Paul Keim, a microbiologist at Northern Arizona University in Flagstaff. All current board members are to be replaced.



STONY BROOK UNIV.

EVENTS

## Student sit-in

Students from Nile University in Giza, Egypt, last week forced their way into the Zewail City of Science and Technology on the outskirts of Cairo. They were protesting about the university no longer having access to buildings on the Cairo site. The university had built on land given to it by former-president Hosni Mubarak's government, but that gift was rescinded after the January 2011 revolution and the land given to the Zewail City. Nobel laureate Ahmed Zewail, a chemist at the California Institute of Technology in Pasadena, is leading the negotiations with Nile University to try to settle the dispute. See go.nature.com/juxrba for more.

## COMING UP

**6–8 SEPTEMBER**
The progress of projects to chart the epigenome will be reviewed at the International Human Epigenome Consortium's second meeting in Seoul, South Korea.
**go.nature.com/znccst**

**10 SEPTEMBER**
The Balzan prizes are announced in Milan, Italy. This year sees two awards set aside for the sciences: for epigenetics and solid-Earth sciences (each worth US$787,000).
**go.nature.com/wfww6q**

## Virus puzzles

An outbreak of hantavirus originating in California's Yosemite National Park has so far affected at least six people, two of whom have died, the National Park Service (NPS) announced on 31 August. The NPS has tried to contact around 1,700 people who stayed in cabins at one of the park villages between mid-June and mid-August. The virus is spread by rodent droppings, and this outbreak has puzzled medical researchers as the rare previous cases originated from a single cabin on each occasion. But this time, the infected visitors had stayed in different cabins. See go.nature.com/v86lxo for more.

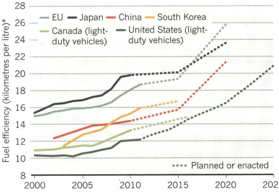
**➲ NATURE.COM**
**For daily news updates see:**
**www.nature.com/news**

## TREND WATCH

US President Barack Obama has signed new rules requiring car and truck manufacturers to almost double average fuel efficiency by 2025. First announced a year ago, the standards approved on 28 August would see US cars reach the current efficiency of Japanese cars by the mid-2020s (see chart). They would also bring emissions down to around 107 grams of carbon dioxide per kilometre travelled (behind the target of 95 g $CO_2$ per km set by the European Commission for 2020).

### CRACKING DOWN ON GAS GUZZLERS

Finalized US vehicle standards would almost double fuel efficiency by 2025 — but would still lag behind other nations.



*Normalized to Corporate Average Fuel Efficiency (CAFE) standards.

SOURCE: ICCT

**A?**
**Aalto University**

Finland is one of the top performing countries in the world in terms of education, innovation and the quality of life. A flagship project of the recent university reform in Finland, Aalto University was established 2010 by a merger of three Finnish universities: Helsinki School of Economics, Helsinki University of Technology and the University of Art and Design Helsinki. Our goal is to become a world class university by the year 2020, inspiring innovation and entrepreneurship by combining art and science with technology and business (www.aalto.fi/en/about/strategy).

**Aalto University is now seeking to appoint three**

# *Deans*

**to lead the operations of the**

School of ***Chemical Technology*** (chem.aalto.fi/en),
School of ***Engineering*** (eng.aalto.fi/en) and
School of ***Science*** (sci.aalto.fi/en), respectively.

The Deans report directly to the University President and are members of the senior leadership team of the University. The Dean has overall responsibility for directing the school to achieve our world-class targets. He/she will be leading the development and implementation of the school's strategy and is responsible for day-to-day management of the school within the framework of the Aalto University strategy and policies. The Dean has a central role in setting school-level priorities for research and education as well as deciding on the allocation of resources for the academic departments. A key task for the Dean is the recruitment of new faculty according to the recently established Tenure Track career system.

**A successful candidate will have**

- a doctoral degree and proven track record of research and education excellence
- strong, international academic background, and experience in multidisciplinary research
- outstanding strategic capability to drive change by promoting creative solutions and by inspiring people
- ability to develop, implement and promote institutional objectives
- experience of leading academic communities, and evaluating and mentoring faculty
- excellent interpersonal skills to support effective communication between all stakeholders
- fluency in English, ability to speak Finnish or Swedish is an advantage

The Deans are appointed by the Board of the University for a five-year term. Salary is negotiable depending on qualifications and experience. Aalto University is an equal opportunity employer, particularly striving to achieve a more equal gender balance in our university.

**Furher information:**
Ilkka Niemelä, Deputy President,
Tel. +358 50 511 3013, e-mail ilkka.niemela@aalto.fi

**How to Apply:**
Complete application package should include the following:

- Application Letter & Resume / CV
- Copy of transcript supporting the highest earned degree
- At least 3 professional letters of reference

Completed application package should be preferably e-mailed to registry@aalto.fi in pdf format. It is also possible to mail application to postal address Aalto University Otaniemi Registry Office, PO Box 11000, 00076 AALTO, Finland.

Applications should be submitted by Friday, September 28, 2012.

W225868R

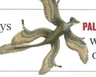
# NEWS IN FOCUS

**CONSERVATION** Dissenter takes an axe to India's forest claims **p.14**

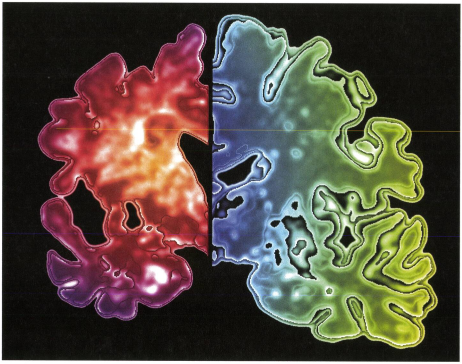
**DATABASES** US funding cuts could leave big data homeless **p.19**

**SPACE** Voyager probe says a long goodbye to the Solar System **p.20**

**PALAEONTOLOGY** Flying high with China's feathered dinosaurs **p.22**



PASIEKA/SPL



Amyloid plaques accumulate in the brains of Alzheimer's patients (left), but not in unaffected brains (right).

MEDICAL RESEARCH

# Alzheimer's drugs take a new tack

*Hopes pinned on pre-emptive clinical trials after latest setbacks.*

BY EWEN CALLAWAY

After a summer marred by disappointing clinical-trial results in patients with Alzheimer's disease, drug developers are regrouping to plot a fresh course in the battle against the devastating disorder.

The bad news began in July and August, when Johnson & Johnson and Pfizer learned that their biological drug bapineuzumab had failed to show any benefit in two large trials. Then, on 24 August, Eli Lilly said that its drug solanezumab had not hit its goal of significantly slowing the memory decline and dementia that characterize Alzheimer's disease.

Both of the failed drugs targeted amyloid-β, a protein that forms plaques in the brains of patients with the disease and that has long been the prime suspect for causing it. But rather than abandoning the amyloid hypothesis, scientists are pinning their hopes on innovative clinical-trial designs and new diagnostics that would allow them to test compounds earlier in the disease and gauge their efficacy more quickly.

Many worry, however, that investors spooked by the hundreds of millions of dollars spent on failed trials will be reluctant to support a continuing search for effective treatments for Alzheimer's and other dementias, which affect an estimated 36 million people worldwide. "Money is tight," says Husseini Manji, global therapeutic area head in neuroscience at Johnson & Johnson in New Brunswick, New Jersey. But "we're still very committed. We think this is a major societal problem that needs tackling."

➲ **NATURE.COM**
Read *Nature*'s Outlook on Alzheimer's disease:
go.nature.com/hdiuds

Amyloid-β plaques are thought to cause Alzheimer's disease by killing neurons and severing their connections to their neighbours. But the evidence is circumstantial. Autopsies of patients show that larger numbers of plaques occur in more severe cases of the disease. Also, mutations in the gene responsible for amyloid-β seem to have either a risk-enhancing or a protective effect. Yet despite all the money invested in amyloid-targeting drugs, "we need to confirm or refute the amyloid hypothesis", says Paul Aisen, a neuroscientist at the University of California, San Diego.

The first results for solanezumab, released by Eli Lilly, which is headquartered in Indianapolis, Indiana, seem to support the hypothesis. The drug is meant to recognize and block amyloid-β before it forms plaques. In patients with mild and moderate forms of disease, however, solanezumab failed to meet its main goals of slowing the decline in memory and other cognitive measures, or in the ability to perform tasks such as eating and maintaining personal care. But other analyses suggest that the drug slowed cognitive decline in patients with milder forms of Alzheimer's. No data have been released on the magnitude of these improvements, though, so it is unclear whether they are enough to make a difference to patients' lives.

"From a purely scientific standpoint, we're pleased at the results," says Eric Siemers, medical director of Lilly's Alzheimer's team. "These are the first clinical-trial data that would also support the amyloid hypothesis." Investors and scientists will get a clearer picture this autumn, when further data from this summer's trials of more than 2,000 patients will be presented at conferences.

The bapineuzumab trials seem to have been more of an unqualified failure. This antibody drug targets the amyloid-β plaques, in hopes of awakening the immune system to clear them from the brain. But two trials in approximately 2,400 patients failed to show any benefit ▶

**TACKLING ALZHEIMER'S EARLY**
Three studies aim to assess the effects of trial drugs on asymptomatic people.

| Trial name | Aim | Length | Size | Cost |
|---|---|---|---|---|
| Alzheimer's Prevention Initiative | To test crenezumab in people who have mutations in the presenilin 1 gene and other genes that cause Alzheimer's in middle age. | 5 years | ~300 people | $100 million |
| Dominantly Inherited Alzheimer Network | To test three drugs on asymptomatic people with Alzheimer's-linked mutations in genes for presenilins 1 and 2, and amyloid precursor protein. | 5 years | 160 people | $60 million for 2 years |
| Anti-amyloid treatment in asymptomatic Alzheimer's disease | To test a drug in asymptomatic people who have high levels of amyloid-β, and some who have a gene variant that increases their risk of Alzheimer's. | 3 years | 1,000 people | $110 million |

▸ compared with a placebo, although this may have been because the drug was administered in lower doses than solanezumab, owing to its higher toxicity. Johnson & Johnson and its partner Pfizer, headquartered in New York city, say that they will vastly scale back development of bapineuzumab.

Increasingly, researchers think that the problem lies not so much with the strategy of targeting amyloid-β as with the timing of treatment. "The major conundrum in the field is: 'are we just treating people too late?'," says Ronald Petersen, director of the Alzheimer's Disease Research Center at the Mayo Clinic in Rochester, Minnesota. Like the fatty plaques in coronary arteries, amyloid-β plaques accrue over a lifetime, says Petersen. And so, just as cholesterol-lowering statins are prescribed for patients in middle age to stave off heart disease in later life, amyloid-blocking drugs given in middle age may prevent Alzheimer's, Petersen says.

But no one knows when amyloid-blocking drugs would need to be taken to prevent the disease, and researchers might have to track tens of thousands of people for decades to determine whether a preventive drug worked. "You can't take every 30-year-old off the street and try a prevention study," says Manji.

Nonetheless, three studies are set to begin by next year that will test whether anti-amyloid drugs can forestall early symptoms of Alzheimer's and arrest cognitive decline in patients who, on the basis of genetic predisposition or amyloid levels, have been identified as being at increased risk of developing the disease (see 'Tackling Alzheimer's early').

The Alzheimer's Prevention Initiative will test crenezumab, a drug developed by Genentech, based in South San Francisco, California, in a large Colombian family that has a rare mutation predisposing members to develop Alzheimer's in middle age. The US$100-million trial will focus on asymptomatic family members for up to five years to see if the drug can stave off their inevitable cognitive decline. The trial will also seek to identify biomarkers, such as amyloid levels from brain scans and in cerebrospinal fluid, that could be used to assess whether crenezumab and other drugs are effective.

> *"The major conundrum in the field is: 'are we just treating people too late?'."*

"We need to launch a new era in Alzheimer's-prevention research to set the stage to rapidly evaluate treatments," says Eric Reiman, executive director of Banner Alzheimer's Institute in Phoenix, Arizona, who is co-leading the Colombia trial. With such markers identified, drug companies could quickly get a sense of whether or not a drug is preventing Alzheimer's, saving precious money and time, he says.

Drug agencies, including the US Food and Drug Administration and the European Medicines Agency, are keeping a close watch on those efforts. In theory, approval for preventive drugs could be assessed on the basis of clinical trials measuring changes in biomarkers, or surrogates, instead of traditional measures of cognitive improvement. However, regulatory agencies are likely to set a very high bar for what constitutes a proven surrogate, says Siemers.

Reiman's study is already bankrolled. But the two other imminent trials — one led by the Alzheimer's Disease Cooperative Study, a US government-funded programme, and the other by researchers at Washington University School of Medicine in St Louis, Missouri — are still looking for money. Many Alzheimer's experts hope that this summer's bleak news will not scare off investors.

"We've had this concern for quite some time," says Reiman, "that if these trials were negative we would see some major stakeholders and investors abandon amyloid-modifying treatments. We think that would be throwing the baby out with the bath water, and abandoning Alzheimer's disease." ■

---

CONSERVATION

# India's forest area in doubt

*Reliance on satellite data blamed for over-optimistic estimates of forest cover.*

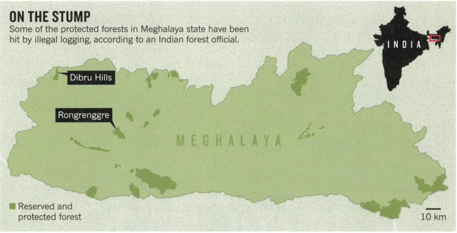
**BY NATASHA GILBERT**

To judge from India's official surveys, the protection of its forests is a success. Somehow, this resource-hungry country of 1.2 billion people is managing to preserve its rich forests almost intact in the face of growing demands for timber and agricultural land.

But a senior official responsible for assessing the health of the nation's forests says that recent surveys have overestimated the extent of the remaining forests. Ranjit Gill of the Forest Survey of India (FSI) claims that illegal felling of valuable teak and sal trees has devastated supposedly protected forests in the northeast of the country. He and other experts also say that an over-reliance on inadequate imaging by an Indian satellite system is making such destruction easy to overlook.

In February, the FSI, part of the government's Ministry of Environment and Forests, released the *India State of Forest Report 2011*. This biennial survey used images from India's remote-sensing satellite system and estimated that forest covered 692,027 square kilometres of the country — roughly 23% of India's land area — a decline of just 367 km$^2$ on the tally reported in 2009, and a much smaller loss than in Brazil, for example, where more than 13,000 km$^2$ of forest was cleared over the same period. But Gill, a joint director of the FSI, is openly critical of the FSI's assessment.

"We have to accept the grave reality that the current figure of forest cover in India is way over the top and based on facile assumptions," Gill argues. To bring these allegations to light,



**ON THE STUMP**

Some of the protected forests in Meghalaya state have been hit by illegal logging, according to an Indian forest official.

FOREST & ENVIRON. DEPT, MEGHALAYA

RANJIT SINGH GILL

Forest officer Ranjit Gill says that he has evidence of widespread deforestation in Meghalaya (above).

he has mounted a legal case for consideration by India's Central Empowered Committee (CEC), a panel of experts appointed by the nation's Supreme Court to rule on issues concerning forests and wildlife.

Gill alleges that the government of Meghalaya state in northeast India has failed to act sufficiently on evidence that illegal felling and coal mining is ravaging the region's protected forests. He says that he has seen the deforested areas at first-hand, and reported them to the state government (see 'On the stump'). He is also concerned that the 2011 forest report records large areas in Meghalaya as open or dense forest, when he believes that much of the land had been cleared and then allowed to regrow saplings or bamboo.

On a field survey last year, Gill and three FSI colleagues saw that parts of the Dibru Hills protected forest in Meghalaya had been illegally felled. He confirmed his field observations with 2006 data from the LANDSAT Earth-observing satellites operated by NASA and the US Geological Survey. The satellite data showed that roughly 150,000 trees in the area had been cut down in the preceding years, across an area of about 10 km$^2$.

Gill also points to an investigation in 2006 by Meghalaya state's forest and environment department. The report, which he obtained through a freedom-of-information request and showed to *Nature*, found illegal saw mills operating in the area, as well as freshly felled logs. The region has "come under tremendous pressure and suffered serious depletion, which has reached alarming proportions", that report says.

According to documents submitted to the CEC, the Meghalaya state government claims that only 670 trees were felled in the Dibru Hills forest from 2004 to 2007. In Gill's view, "the records and reports of the government of Meghalaya are not a true picture of the positions on the ground". P. B. O. Warjri, chief secretary of the government of Meghalaya, told *Nature* that Gill's claims are "not true".

But another state government report obtained by Gill documents similar illegal deforestation in the nearby Rongrenggre protected forest, where 60–70% of the tree cover has been lost. The report also found evidence that local forest rangers were involved in the illegal timber trade, and that illegal coal mining in the area was taking place in "full knowledge" of the rangers. Gill is concerned that similar lapses are happening, and not being reported, in other parts of the country.

Other tropical-forest researchers share Gill's fears about India's forests. "The ongoing loss and attrition of native forest in India is quite widespread, although it isn't being captured by the government's satellite data on forest cover," says William Laurance, a conservation biologist at James Cook University in Cairns, Queensland, Australia. "Much of this forest disruption is illegal, and encroachment into protected areas and reserves is not uncommon, in my experience."

**➲ NATURE.COM**
Read more at Nature India:
www.nature.com/nindia

Anil Kumar Wahal, the director of the FSI, denies that forest cover has been overestimated. The FSI team that conducted the field visit in May 2011, of which Gill was part, "reported a few sporadic patches of felling, and old stumps in the field, but nothing as glaring as felling of vast swathes of forest", he says. But Wahal admits that the "selective" cutting of trees "would not register in the satellite imagery due to the technological limitation of the medium-resolution sensor used for the purpose of forest-cover mapping".

Gill notes that the instrument, which flies on an Indian remote-sensing satellite, produces images with a resolution of 23.5 metres per pixel, too coarse to unequivocally identify small-scale deforestation. Instead, he says, the forest survey should use a newer instrument, already operating on an Indian satellite, that provides a resolution of 5.8 metres per pixel.

The FSI uses the lower-resolution instrument for its national survey because it offers continuous coverage of very large areas, explains Wahal. "Gap-free data are really essential," he says. "Using high-resolution data would also entail much more manpower and time, so a balance has to be struck." The FSI is, however, using the higher-resolution instrument for some small-scale surveys, he adds.

Gill argues that the FSI still needs to conduct more on-the-ground surveys to corroborate its satellite estimates of forest cover. Without this reality check, it can be difficult to tell the difference between native forests and, for example, bamboo. He is calling on the CEC to order a visit to the forests to investigate the extent of the destruction. A verdict is expected from the CEC by the end of the year.

Last year, India's government grabbed headlines with a US$10-billion, decade-long plan — the National Mission for a Green India — to create or improve 10 million hectares of forest. But if Gill is right, it faces a more urgent task: to chart and protect the forests that remain. ■

TREAT PEOPLE RIGHT NOW.
PEOPLE OVER PHARMA
COMMUNITY POWER
TRADE DEAL KILLS
TA: DEATH UNDER PATEN

PUBLIC CITIZEN/GLOBAL TRADE WATCH

**Trade rules that would raise the cost of HIV medicines come under fire at a July rally in Washington DC.**

PUBLIC HEALTH

# Trade deal to curb generic-drug use

*Tighter patent rules could raise drug costs in poor countries.*

**BY AMY MAXMEN**

"Wanted," the notice reads, in an American old-west style font, "Negotiating text of the Trans-Pacific Partnership Agreement." The online advert invites visitors to contribute to a reward payable to the WikiLeaks website should it manage to expose the trade agreement. As *Nature* went to press, the reward stood at US$24,490.

The tactic, employed by the activist group Just Foreign Policy in Washington DC, may be extreme, but it reflects a broader unease over a negotiation process that the advert says "could affect the health and welfare of billions of people". At issue are industry-friendly rules governing drug patents that could be written into the final text of the Trans-Pacific Partnership Agreement (TPP). The provisions could boost drug development and profits for the pharmaceutical industry, but also curb the use of cheaper generic medicines in low- and middle-income nations.

"In many parts of the world, access to generic drugs means the difference between life and death," says US congressman Henry Waxman (Democrat, California). He is one of several US politicians voicing concern over the closed-door TPP negotiations and the influence that the pharmaceutical industry is thought to be exerting on the process through US trade representatives. With the latest round of talks set to begin on 6 September in Leesburg, Virginia, public-health advocates are expressing fears that the outcome will reduce access to medicines.

Besides the United States, ten Pacific countries representing 34% of US trade have so far agreed to join the TPP — Australia, New Zealand, Singapore, Malaysia, Brunei, Vietnam, Peru, Chile, Canada and Mexico. The agreement, which could come into effect as early as next year, spans several trade areas, meaning that some countries may be tempted to forgo access to generic drugs in exchange for better access to US markets in other industries.

According to previously leaked documents, the TPP looks likely to strengthen patent protection for drugs more than any trade agreement so far. Whereas the current World Trade Organization (WTO) agreement sets a minimum 20-year period for patents around the world, the TPP would follow US practice in extending patents beyond 20 years when the drug-approval process has delayed a drug's market entrance. Partner countries would also be pressed to award new patents for off-patent drugs that have been formulated in a new way or approved for a new set of patients.

This practice restricts access to medicines in poor countries because it extends patent monopolies. For example, according to Médecins Sans Frontières (also known as Doctors Without Borders) in Geneva, Switzerland, countries that have rejected patents on new formulations of the off-patent HIV drug Abacavir now sell generic versions for as little as $139 per person per year, whereas in Malaysia paediatric Abacavir costs $1,200 per child per year, because the country granted the new formulation a patent. But a spokesperson from the Office of the US Trade Representative says that patenting new formulations of old drugs provides an incentive for drug companies to develop adaptations "that are valued in developing countries, like heat-stabilized medicines for places without refrigeration".

Industry stakeholders say that drug companies need greater protection as the industry enters an unprecedented period of patent expirations (see *Nature* **480,** 16–17; 2011) and faces stiff competition from generics produced in India and China. They argue that sales of generics need to be restricted if companies are to recoup the millions they invest in developing new drugs. "If TPP countries wish to be those in which innovation flourishes, they should have strong intellectual property," says Stephen Ezell, senior analyst at the Information Technology and Innovation Foundation, a non-profit think tank in Washington DC that supports patent extensions.

The negotiators are considering special protections for biologic drugs — those based on large biological molecules. One possibility under discussion would grant companies a 12-year period of exclusivity on clinical-trial data related to the biologics they develop. Makers of equivalents of small-molecule drugs rely on such data when they seek government approval for their products. Without access to the data, the generics company would have to repeat the costly clinical trials or delay the time-consuming approval process for its product by 12 years. Charlene Barshefsky, a former US trade representative who now advises companies on trade law, explains that the biologics market, which was worth US$149 billion globally in 2010, needs extra protection because biologics cost more to develop than small-molecule drugs. "I am not saying that a foreign innovator cannot develop their own biologic drug, they just need to do their own homework," she says.

More generally, stronger patent provisions would harm small, domestic manufacturers of generic drugs in Malaysia and Vietnam, says Shawn Brown, formerly vice-president for international affairs and state government at the Generic Pharmaceutical Association based in Washington DC. They would also cut sales for larger generics manufacturers in the United

States, Australia and Canada that supply low-cost drugs to the world.

Some countries whose governments purchase drugs with a set budget are also alarmed by signs that the TPP may grant new negotiating powers to the industry. In New Zealand, for example, a government agency called Pharmac determines whether the benefits of a new drug warrant the cost, or if the country is better off sticking with a cheaper alternative. A leaked TPP provision would empower drug companies to appeal such decisions. "We have good processes for ensuring what is for the good of our population, not for the good of lobby groups, and I don't see why they need to interfere with that," says Marilyn Head, a policy analyst at the New Zealand Nurses Organisation in Wellington, who adds: "Bugger off, quite frankly." ■

---

CHEMISTRY

# Electro-optic dye triggers ethics row

## *Dispute puts focus on reporting standards for major grants.*

BY EUGENIE SAMUEL REICH

When a colleague questions a researcher's hypothesis, how far must the researcher go in telling his prospective funders about those doubts?

The question sits at the heart of a dispute that has prompted a government review of alleged omissions in reports from a science and technology centre funded by grants totalling US$36 million over 10 years from the National Science Foundation (NSF). The review, by the NSF's inspector general, is not yet complete, but the affair highlights a grey area in the agency's rules for grant recipients: although the rules require principal investigators to disclose any problems they encounter in pursuit of their research goals, they offer no guidance on how to assess when a colleague's scepticism about a specific issue merits reporting.

The issue became public in late July, when Bart Kahr, a chemist at New York University in New York city, described his side of the dispute at a meeting of the American Crystallographic Association in Boston, Massachusetts. But it goes back more than a decade, to work led by Larry Dalton at the University of Washington in Seattle in 2000. Motivated by the rapid expansion of the Internet, the group was developing modulators, colloquially called 'opto-chips', that convert electrical to optical signals, a more efficient medium for long-distance communication. Dalton and his team reported[1] record-breaking performances by electro-optic devices based on dye molecules they had designed. And their paper suggested that the key to the devices' performance lay in the way the molecules lined up in an electric field.

The result was discussed in a 2001 grant proposal to the NSF, which subsequently funded the Center on Materials and Devices for Information Technology Research at the University of Washington, with Dalton as its director. Research continued on the devices, and Kahr joined the centre in 2003. Several groups at the centre and elsewhere were continuing to report improved performances for the devices, but Kahr began to doubt the mechanism that had been proposed to explain how they worked.

Kahr obtained samples of dye molecules from another researcher at the centre, Alex Jen, and measured their absorption of polarized light — a way to test their alignment — in an electric field. Kahr reported to Jen that his results suggested there was no strong alignment and that future efforts to improve the devices by optimizing the dye alignment might not work unless the mechanism was understood. But the centre's annual report to the NSF for 2003–04 did not mention Kahr's findings. Jen, who wrote the relevant section, explains that he had a wealth of material to include, and that there was no effort to omit Kahr's results because they challenged an aspect of the centre's research direction.

Alarmed at what he regarded as an unethical omission, Kahr complained in 2004 to chemist Alvin Kwiram, then the centre's executive director. Kwiram says that Kahr's doubts were a distraction from the centre's main goal, which was to build and improve working devices. Although Kahr believed that understanding the mechanism was necessary to improve the devices as quickly as possible, Kwiram and others felt that they were already being made more effective even though the mechanism was in dispute. "This issue [of the mechanism] was like a mosquito buzzing around and it was like don't bite me right now when we've got bigger fish to fry," Kwiram says.

The centre submitted two more annual reports without mentioning Kahr's finding that the alignment was weak, and in 2006 the centre's grant came up for a five-year renewal. Phil Reid, a chemist at the centre who is now its director, says that during a site visit by NSF reviewers, Jen mentioned theoretical work suggesting that the dye molecules might not be aligned as strongly as supposed — work also mentioned in the 2005–06 annual report although not in connection with Kahr and his concerns. Kahr says that he did not have an opportunity to present his data to the NSF reviewers, and that he subsequently lost funding he had been receiving through the centre.

Kahr moved to New York University in 2009. In 2011, Reid, Jen, Dalton and Bruce Robinson, a theoretical chemist at the University of Washington, published a paper[2] presenting their own evidence that some dye molecules similar to those used in the original work align only weakly in an electric field — findings that paralleled those of Kahr. Robinson sees this simply as the resolution of a scientific disagreement, not a matter of research ethics. "Bart was right," says Robinson, "but so what?"

After receiving copies of Kahr's e-mails to centre members raising ethical concerns about the omissions, the University of Washington's Office of Scholarly Integrity and Ana Mari Cauce, dean of the university's College of Arts and Sciences at the time, conducted separate investigations of his allegations in 2010 and 2011. Both cleared Dalton and Jen — the only targets of Kahr's accusations — of any violation of ethics. Cauce, who is now the university's provost, explained in a letter to Kahr that Jen's omission of Kahr's data from the annual reports was justified because the data were preliminary and because there was a scientific disagreement about whether the molecules were aligned.

**NSF rules offer no guidance on how to assess when a colleague's scepticism merits reporting.**

But Kahr remained unsatisfied and in January 2011 submitted allegations to the NSF's Office of Inspector General. Susan Carnohan, a spokeswoman for the inspector general, told *Nature* that the office does not comment on ongoing investigations.

Jason Borenstein, a philosopher who teaches responsible conduct of research to science and engineering students at Georgia Institute of Technology in Atlanta, believes that grant applicants should generally disclose a colleague's doubts in their reports to funders. "Typically it is preferred, if there is space, to say there is another viewpoint that could be presented but we believe ours is right for the following reasons," he says. "That will make a better case to the grant reviewers." ■

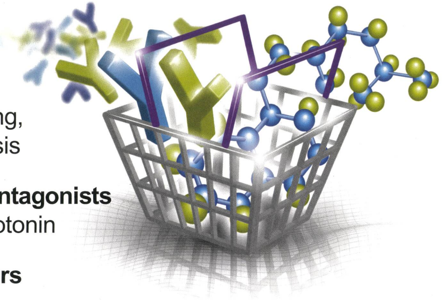
1. Shi, Y. *et al. Science* **288,** 119–122 (2000).
2. Olbricht, B. C. *et al. J. Phys. Chem. B* **115,** 231–241 (2011).

**All your research tools in one basket**

**abcam**Biochemicals®

**Discover more with high purity standard, novel & exclusive biochemicals:**

- **Enzyme inhibitors**
  Akt, MAPK, Caspases
- **Signaling tools**
  Apoptosis, $Ca^{2+}$ signaling, Endocytosis & exocytosis
- **Receptor agonists & antagonists**
  Glutamate, GABA, Serotonin
- **Ion channel modulators**
  $Ca^{2+}$, $K^+$, $Na^+$



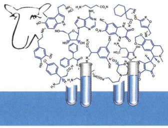
**What to expect when you order Abcam Biochemicals:**

✓ High purity and exceptional quality guaranteed

✓ Competitively priced

✓ Excellent supporting data

✓ Friendly and knowledgeable customer service



**Discover more** at abcam.com/biochem

INFORMATICS

# Databases fight funding cuts

*Online tools are becoming ever more important to biology, but financial support is unstable.*

**BY MONYA BAKER**

**ENDANGERED DATABASES**
The US National Library of Medicine (NLM) is cutting resources that biologists say are vital to their research.

| Resource | NLM-funded since | Function | Usage | Last NLM award |
|---|---|---|---|---|
| Protégé | 1990 | Creating tools to organize and analyse data | 200,000 registered users | $956,625 |
| BioMagResBank | 1990 | Holds spectroscopy data for biomolecules | 500–1,000 unique users per day | $727,129 |
| Repbase | 1994 | Identifying families of non-coding DNA across species | 8,000 registered users | $551,544 |
| REBASE | 1995 | Finding where enzymes bind to and cut DNA | 495,844 website hits per month | $235,911 |
| CASP | 2001 | Testing techniques to predict protein structure | More than 100 research groups participate | $515,168 |

In the era of 'big data', it is a bitter blow for scientists to lose access to the online tools they use to analyse and share terabytes of information. Yet funding cuts by the US National Library of Medicine (NLM) are threatening five widely used biological databases, and user communities are now rallying to save them. "The idea that this resource could just disappear is a serious problem for everyone who relies on it," says Mark Musen, a bioinformatician at Stanford University in California, and manager of Protégé, which provides open-source software to organize and interrelate biological data.

Protégé has 200,000 registered users, and the NLM, part of the National Institutes of Health (NIH) in Bethesda, Maryland, has contributed millions of dollars to maintain it. But in 2007, the NLM decided that it would stop supporting infrastructure grants and would redirect resources to informatics research, says Valerie Florance, director of extramural programmes at the library. Consequently, the NLM's support for Protégé and similar projects is not being renewed (see 'Endangered databases'). "It is not a reflection of the value of the resources to any of their users," says Florance. "It is part of our determination to put our funds into research and training."

The argument is playing out at other funding agencies, says David Botstein, a genomicist at Princeton University in New Jersey, and a member of the NIH Data and Informatics Working Group, which published a draft report on the issue in June. "The whole system is rigged against infrastructure of any kind," he says, predicting that "many, many resources" will face similar funding crises in the near future.

The Biological Magnetic Resonance Data Bank (BioMagResBank, or BMRB), for example, has been funded by the NLM since 1990 and holds more than 7,500 entries on biomolecules. Structural biologists use the nuclear magnetic resonance data to probe questions such as how proteins contort as they catalyse reactions.

More than 90 scientists have written letters to *Nature Structural and Molecular Biology* this month in support of the BMRB (J. Markley *et al. Nature Struct. Molec. Biol.* **19,** 854–860; 2012). Inês Chen, chief editor of the journal, says that losing the database would deprive researchers of access to crucial data. "As journals, we cannot host all the data that are part of the paper, and so if they disappear, it's a big deal."

John Markley, director of the BMRB and a structural biologist at the University of Wisconsin-Madison, hopes to attract other federal funders to support the database.

Another option is to charge users, but Musen calls that "absurd", arguing that it would discourage scientists from accessing sites and, in the case of Protégé, from contributing the code and plug-ins that make it a useful resource. Musen wants to win funding from the NIH to keep Protégé going as a key component of new research projects. In June 2011, he submitted a grant application with more than 100 letters of support from scientists; reviewers acknowledged the letters but said that they had nothing to do with the grant's specific research goals, and turned it down. Musen resubmitted the application, and should learn the results this month.

Other databases are putting their trust in commercial sponsors. REBASE, which holds data on where enzymes bind to and cut DNA, is partially supported by laboratory-reagent company New England Biolabs of Ipswich, Massachusetts. When federal money runs out in 2014, the company will take on the full costs, says Richard Roberts, chief scientific officer of New England Biolabs and founder of REBASE. But he acknowledges that this potentially leaves the database at the mercy of shifting commercial priorities.

The least vulnerable databases are those directly run by government agencies, says Francis Ouellette, a bioinformatician at the Ontario Institute for Cancer Research in Toronto, Canada. Investigator-driven databases face more challenges because "they don't fit the research-based standard model" used to dispense grants. Cutting funding for poorly performing or obsolete databases is sensible, says Ouellette, but choking established sites that have significant user communities is "really short-sighted. If it's a good database it should be maintained."

Florance argues that the NLM should back innovation, which is difficult when its funds are tied up in infrastructure. "I don't think anyone would say that because they got a grant and built a database, they should get money forever."

One solution, says Musen, could be to wean successful projects off investigator-initiated grants and move them into the NIH's longer-term intramural programmes. But Botstein thinks that would require a philosophical change at the agency. "What's really required is an understanding of the larger problem," he says. "This is a big thing, and it will be a big thing for years to come." ■

**MORE ONLINE**

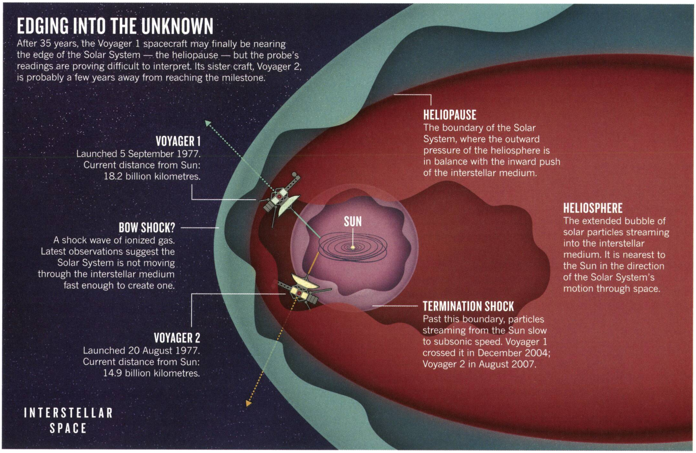
**Q&A**

IAU

Astronomer Cui Xiangqun on China's plans for more telescopes and probes **go.nature.com/zzhveu**

**NEWS**

- The TB test you can do at home **go.nature.com/wmtneq**
- Controversy over widely used intravenous drips **go.nature.com/eqlej2**
- Ancient humans interbred with Denisovans **go.nature.com/jj8dxt**



## EDGING INTO THE UNKNOWN

After 35 years, the Voyager 1 spacecraft may finally be nearing the edge of the Solar System — the heliopause — but the probe's readings are proving difficult to interpret. Its sister craft, Voyager 2, is probably a few years away from reaching the milestone.

**VOYAGER 1**
Launched 5 September 1977.
Current distance from Sun:
18.2 billion kilometres.

**BOW SHOCK?**
A shock wave of ionized gas. Latest observations suggest the Solar System is not moving through the interstellar medium fast enough to create one.

**VOYAGER 2**
Launched 20 August 1977.
Current distance from Sun:
14.9 billion kilometres.

**HELIOPAUSE**
The boundary of the Solar System, where the outward pressure of the heliosphere is in balance with the inward push of the interstellar medium.

**HELIOSPHERE**
The extended bubble of solar particles streaming into the interstellar medium. It is nearest to the Sun in the direction of the Solar System's motion through space.

**TERMINATION SHOCK**
Past this boundary, particles streaming from the Sun slow to subsonic speed. Voyager 1 crossed it in December 2004; Voyager 2 in August 2007.

ASTROPHYSICS

# Voyager's long goodbye

*NASA probes find surprises at the edge of the Solar System.*

**BY RON COWEN**

Are we there yet? Ed Stone, the project scientist for NASA's two Voyager spacecraft, wants to know. Since their launch in 1977, the probes have ventured billions of kilometres beyond the outer planets. Now, Stone and his colleagues are looking for signs that Voyager 1 may finally be nearing the edge of the Solar System — where the heliosphere, the bubble of electrically charged particles blown outwards by the Sun, gives way to interstellar space (see 'Edging into the unknown').

Detecting and characterizing this threshold — called the heliopause — would be the ultimate bonus for a probe that logged its 35th year in space on 5 September. When Voyager 1 set out, says Stone, a physicist at the California Institute of Technology in Pasadena, who has coordinated the mission since the probes launched, "the space age was only 20 years old and there was no evidence that any spacecraft could travel this long and this far from the Sun".

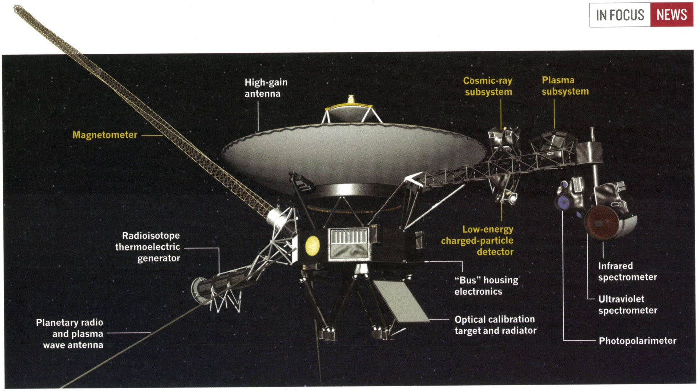
The extraordinarily long-lived Voyager 1 began detecting hints of a boundary region eight years ago. But exiting the Solar System is proving to be a longer and more complicated affair than Stone and his colleagues had anticipated. By the time Voyager 1 is well and truly out, it may have transformed researchers' ideas about the Solar System's invisible edge.

In the latest twist in the story, the craft seems to be traversing an unexpected 'dead zone'. This week, Robert Decker, a space scientist at the Johns Hopkins University Applied Physics Laboratory in Laurel, Maryland, and his colleagues report[1] in *Nature* that at Voyager 1's current location, some 121.6 astronomical units (18.2 billion kilometres) from the Sun, the average velocity of solar particles has dropped to nearly zero. (Voyager 2, which is about 3 billion kilometres closer to the Sun and moving in a different direction, has yet to detect the same reduction in velocity.)

➲ NATURE.COM
Meanwhile, the Curiosity rover explores Mars:
nature.com/curiosity

Decker's team first reported[2] the change last year, when it had measurements of the particles' velocity only in the radial direction, outwards from the Sun. At the time, the team thought that the change was a sign that the craft was nearing the heliopause, where solar particles are expected to collide with powerful winds generated by supernovae that exploded some 5 million to 10 million years ago. The collision would force the solar particles to stop moving outwards and push them sideways, like a stream of water hitting a solid surface.

To test the idea, engineers commanded Voyager 1 to roll on its side seven times, so that its instruments could record particle velocities along a line perpendicular to its course. Given that sending a command to Voyager 1 now takes 17 hours, and that the spacecraft's transmitter runs at 23 watts — about as powerful as a refrigerator light bulb — such communication is a feat in itself. The researchers were astonished to find that the particles had zero velocity in this polar direction, too — indicating that they were almost stationary rather than being buffeted by stellar winds. That


High-gain antenna
Cosmic-ray subsystem
Plasma subsystem
Magnetometer
Low-energy charged-particle detector
Radioisotope thermoelectric generator
"Bus" housing electronics
Infrared spectrometer
Ultraviolet spectrometer
Optical calibration target and radiator
Planetary radio and plasma wave antenna
Photopolarimeter

NASA/JPL-CALTECH

**Voyager 1 was launched in 1977. Four of its original instruments (labelled in yellow) are still returning data on conditions at the edge of the Solar System.**

cannot happen at the heliopause, says Decker. "We therefore conclude … that Voyager 1 is not at the present time close to the heliopause, at least in the form that it has been envisioned," the team writes[1].

Decker and his colleagues now think that since 2010, when the craft first recorded a velocity drop, it has been in an antechamber to the heliopause, at least 1 billion kilometres thick. Why the particles are becalmed remains a mystery, says Stamatios Krimigis, a space scientist at Johns Hopkins and a co-author of the paper. This leaves theorists in a bind. "There no longer exists any guidance on what constitutes getting out of the Solar System and into the Galaxy," says Krimigis.

Gary Zank, a theoretical physicist at the University of Alabama in Huntsville, disagrees. "I don't regard the paper as forcing us to revise our models," he says. His team and others theorize[3] that a magnetic wall in the outer heliosphere, caused by a pile-up of magnetic field lines, could slow down the flow of charged particles and account for the near-zero velocities recorded by Voyager 1.

Although the craft has not yet made it to the heliopause, the boundary may be within reach. This May, Voyager 1 recorded unprecedented bursts of cosmic rays — highly energized protons and atomic nuclei — coming from outside the Solar System. The spikes returned in July, this time along with a drop in the incidence of lower-energy cosmic rays thought to be accelerated in the Solar System. The changes suggest that Voyager 1 is nearing the fringe of the Solar System, and could cross the heliopause by the end of the year, says Krimigis. But, he adds, "nature seems to be much more imaginative than we are, so I could be quite wrong".

Indeed, David McComas, a physicist at the Southwest Research Institute in San Antonio, Texas, and Nathan Schwadron, a plasma physicist at the University of New Hampshire in Durham, suggest an alternative explanation. In an article in press in *The Astrophysical Journal*, they propose that Voyager 1 is in a region where magnetic field lines running through the outer heliosphere link up with the magnetic field of the rest of the Galaxy. Here the field would create a conduit for galactic cosmic rays, causing the spikes in detection. Cosmic rays accelerated within the heliosphere would tend to move along other field lines and be less likely to get to Voyager. If this model is correct, say McComas and Schwadron, the heliopause may still be years away.

> ***"There no longer exists any guidance on what constitutes getting out of the Solar System."***

When Voyager 1 does leave the Solar System, it may meet further surprises. Researchers have long assumed that a bow shock lies outside the heliopause. Similar to the shock wave around a supersonic aircraft, the bow shock is thought to form as the Solar System ploughs through the interstellar medium, forcing the local ionized gas to change density abruptly and discontinuously. But in May, McComas and his colleagues reported[4] that data from NASA's Interstellar Boundary Explorer (IBEX) mission cast doubt on this picture. From Earth orbit, IBEX probes the interstellar medium by detecting electrically neutral atoms that slip into the Solar System through the heliopause. Its measurements suggest that the Sun and planets are moving through the interstellar medium about 12% slower than previously calculated — too slow to generate a bow shock.

None of this uncertainty bothers Stone, who expects both Voyagers to cross the heliopause well before 2025, when the craft are due to go silent as the plutonium isotopes that supply their power run out. On the contrary, Stone adds, he is pleased that the one-way journey has taken so many unexpected turns. "One thing Voyager has taught us is to be prepared to be surprised." ■ **SEE EDITORIAL P.6**

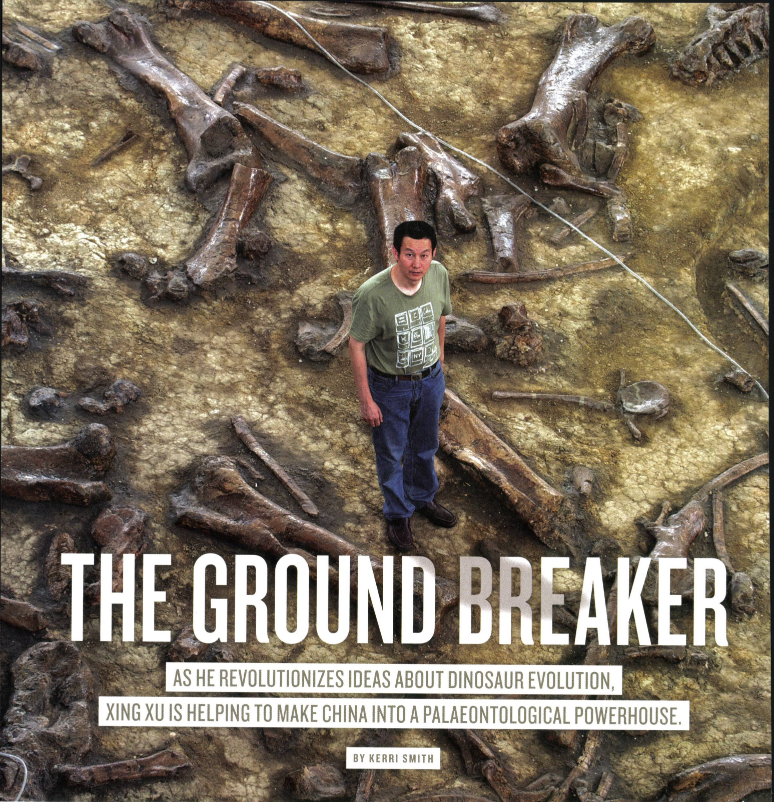
1. Decker, R. B., Krimigis, S. M., Roelof, E. C. & Hill, M. E. *Nature* **489,** 124–127 (2012).
2. Krimigis, S. M., Roelof, E. C., Decker, R. B. & Hill, M. E. *Nature* **474,** 359–361 (2011).
3. Zank, G. P. *Space Sci. Rev.* **89,** 413–688 (1999).
4. McComas, D. J. *et al. Science* **336,** 1291–1293 (2012).

## CORRECTIONS

The News Feature 'Making the links' (*Nature* **488,** 448–450; 2012) misspelt David Lazer's name and wrongly located him. He is at Northeastern University in Boston.

The News Feature 'Man of the desert' (*Nature* **488,** 272–274; 2012) got the details of Kröpelin's 2005 trip wrong. The heavy gunfire heard by the team was caused by Darfur rebels killing 20 Sudanese soldiers (not the other way round).

The News Feature 'Armed resistance' (*Nature* **488,** 576–579; 2012) conflated the Puebla campuses of the University of the Americas and the Monterrey Institute of Technology and Higher Education. The former was home to the first nanotechnology lab in Mexico, the latter was the first institute in Latin America to offer an undergraduate programme in the field and had a false bomb alert last August.



# THE GROUND BREAKER

## AS HE REVOLUTIONIZES IDEAS ABOUT DINOSAUR EVOLUTION, XING XU IS HELPING TO MAKE CHINA INTO A PALAEONTOLOGICAL POWERHOUSE.

BY KERRI SMITH

**Xing Xu stands among the remains of duck-billed dinosaurs in Zhucheng, China.**

Palaeontologist Xing Xu bends low over a beautifully preserved specimen of the ancient bird species *Sapeornis*, entombed in a glass museum cabinet in Shandong Province, China. The bird's spindly legs stretch as if it were about to stride forward, even though the creature has been dead for more than 110 million years. From its chicken-sized body juts a fine neck, a delicate skull and the clear imprint of a long, jaunty tail feather — something never seen before in this species.

*Sapeornis* is one of hundreds of plumed specimens pouring out of fossil beds in China — most notably out of the rock formations in Liaoning Province, northeast of Beijing. Some of the Liaoning fossils are the earliest known birds. Others are feathered dinosaurs,

LOU LINWEI

the group that spawned birds millions of years before the age of *Sapeornis*. Together, they are among the most important finds in dinosaur palaeontology in the past century.

Xu is at the centre of that bonanza. He is "the go-to man in China for anything people want to know about dinosaurs", says Paul Barrett, who studies dinosaurs at the Natural History Museum in London and first met Xu in the 1990s, when both were graduate students. Xu, who is based at the Institute of Vertebrate Paleontology and Paleoanthropology (IVPP) in Beijing, has named 60 species so far — more than any other vertebrate palaeontologist alive today. And he is only 43 years old.

In describing the flock of feathered fossils, Xu has helped to show that birds arose from dinosaurs, ending decades of debate. Along the way, he has shed light on the origins of feathers and flight. And he has bucked 150 years of received wisdom by declaring that the fabled genus *Archaeopteryx* is not the oldest known bird, but rather belonged to a group of dinosaurs removed from the avian line[1]. "He has patience and persistence — and an audacity when scientific evidence calls for it," says Zhe-Xi Luo, who studies fossil mammals at the University of Chicago in Illinois.

Even as he unveils new species at a breakneck pace, Xu is concerned about the future of palaeontology in China and the commercialization of fossils. Many of the feathered fossils from Liaoning are dug up by local farmers tending their fields, who try to sell them to the highest bidder. This fossil 'grey market' — it is technically illegal to sell fossils in China, but the practice continues openly — encourages fakery and causes specimens to disappear into private collections. By cultivating a vast network of contacts at important fossil sites in Liaoning and elsewhere, Xu has laboured to ensure that scientists gain access to the best specimens. It's a job that requires hard work and luck, he says. "When I started my career, I never expected that I would have so many discoveries."

> "MY EXCITEMENT IS PROPORTIONAL TO THE INFORMATION YOU GET. AND THOSE WERE REALLY EXCITING FOSSILS."

### DINO DISNEY

Nobody knows what happened about 80 million years ago near what is now the town of Zhucheng in Shandong Province, but it must have been disastrous. On the outskirts of the city, about an hour's flight south of Beijing, hundreds of bones litter a 300-metre stretch of hillside. Palaeontologists have been finding dinosaurs near Zhucheng for decades, but in 2008 local farmers unearthed a large community of duck-billed dinosaurs and others that had apparently died en masse.

Xu was called in to investigate and he is now studying a possible new species of ceratopsian — herbivorous beaked dinosaurs — recovered from the fossil bed. He is also acting as scientific consultant to local administrators, who want to build a dinosaur theme park in Zhucheng. During a visit to the site in June, Xu had hoped to do research, but he ended up correcting display captions and reading through proposals for the park. "In terms of scale it may be comparable to Disneyland," says Xu, a hint of trepidation in his voice.

Fossils are a thriving business as well as a science in China, and palaeontologists often have to negotiate with local prospectors and directors of museums and tourism bureaux to gain access to fossil sites and specimens. Despite Xu's boyish appearance, he is a dexterous diplomat and has managed to arrange for the most scientifically interesting specimens to cross his desk, wherever they are found.

Thanks to those arrangements, Xu has had a bounty of fossils to work on, particularly from Liaoning. The creatures unearthed there are remarkably well preserved, perhaps because they were entombed quickly during volcanic eruptions and mudslides between 160 million and 120 million years ago. The rocks record fine details including the imprints of feathers, which allowed Xu to determine[2] that a fierce 9-metre-long tyrannosaurid, which he named *Yutyrannus*, had a coat of long feathers (see 'Xing Xu's feathered friends'). One of Xu's favourite Liaoning fossils, *Microraptor*, is one of the smallest known dinosaurs not on the avian line. From the imprint of feathers, Xu and his colleagues concluded[3] that *Microraptor* had four wings — one on each arm and leg — and could probably glide. From other Liaoning specimens, he has established[4] that some feathered dinosaurs slept curled up, just like birds.

When he can find the time, Xu does fieldwork of his own (see 'Dinosaur hunting grounds'). He led teams to three sites this summer. Near the northern Chinese town of Lingwu, the excavations turned up a new sauropod — a dinosaur from the same group as *Diplodocus*. In the autonomous region of Inner Mongolia, the Xu group found a new type of bird and what may be a previously unknown theropod — the dinosaur lineage that led to birds. At another northern site, he uncovered a collection of beaked dinosaurs.

To power this dinosaur-discovery factory, Xu runs a lab of 14 people, including five students, seven preparators who carefully separate the fossils from the surrounding rock, one artist and a photographer. Those resources have been known to induce jealousy in Western palaeontologists. The Natural History Museum in London, for example, has just two full-time preparators for about 20 palaeontology curators and researchers, says Barrett.

➲ **NATURE.COM**
For an interview with Xing Xu and a video, visit:
go.nature.com/olfuvi

Xu didn't set out to be a palaeontologist; in fact, he had no idea what a dinosaur was until he entered university. He was born in the poor Western province of Xinjiang in 1969, a few years after his parents relocated there as part of a Cultural Revolution development initiative in which educated couples were forced to move to rural provinces.

He excelled in school and in 1988 earned a place at Peking University in Beijing, the nation's premier university. Xu wanted to study economics, but at the time students had no choice in their degrees. For reasons that are unclear to him, he was obliged to study palaeontology.

### LATE STARTER

Xu's interest in the subject picked up only when he reached the third year of a master's degree at the IVPP. He was studying two specimens that his adviser, Xijin Zhao, had discovered in the 1960s and 1970s and had not found time to analyse fully. They turned out to be the earliest examples of ceratopsians, pushing the record of this group back by up to 30 million years, from the early Cretaceous period, which started 145 million years ago, to the middle or late Jurassic period[5]. "My excitement [over a fossil] is proportional to the information you get from it," says Xu. "And those were really exciting fossils."

Xu's timing was perfect. While he was working on his master's thesis, the trickle of dinosaur species turning up in China grew to a deluge. Funding for palaeontology was increasing; farmers in Liaoning started recognizing the value of the fossils they sometimes found; and a burst of construction meant that new fossils were being unearthed more frequently. As a budding dinosaur palaeontologist, Xu was well placed to study some of those specimens.

However, fortuitous timing can explain only a portion of Xu's productivity. A large part comes from his legendary work ethic. "If I want to learn something I put all my time into it," says Xu. He currently has more than 20 manuscripts in draft form, including one on the *Sapeornis* specimen from the Shandong Tianyu Museum of Nature. He estimates that there are eight or nine new species among the crop of fossils awaiting publication.

Even away from his office, any spare moment is filled with talk of projects. Outside the Tianyu museum, Xu chats to a colleague about *Microraptor* and — to make an anatomical point — starts drawing a diagram of the creature's feathers in the dust on a nearby car.

Xu has an international outlook that also contributes to his success. From the start of his career, he has done what has not come naturally to many Chinese palaeontologists

— building up a fat book of contacts in the United Kingdom and the United States, and publishing much of his work in English in international journals. Playing to a tougher international audience was "really important for my career", says Xu. Chinese journals, he adds, don't require the same level of critique and peer review as international publications.

Luo says that Xu is one of only a few palaeontologists in China to embrace cladistics — a process for determining evolutionary relationships by analysing the features that groups share. Western researchers and international journals have been using cladistics for more than two decades, but it has been slow to catch on in China.

Within his own country, Xu crosses boundaries between the academic and commercial sectors. For example, he has forged a close relationship with Xiaoting Zheng, the former head of a local state-owned gold mine who is now a keen amateur fossil collector, a budding palaeontologist and director of the Tianyu museum. In his museum, Zheng has accumulated one of the largest assemblages of feathered dinosaur fossils in the world. Over the years, Xu has been teaching him what to look out for in his purchases and has analysed some of the acquisitions. The two make a formidable team.

## FEATHERS FLYING

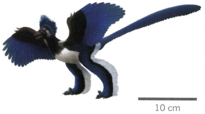
Last year[1], Xu made a big splash with a specimen from the Tianyu museum's collection: a small feathered dinosaur that he named *Xiaotingia zhengi* to honour Zheng. The creature had a shallow snout, a distinctive skull shape and other features that led Xu and his colleagues to place it as a close relative of *Archaeopteryx*. That animal has long been regarded as the oldest known bird, but Xu and his colleagues performed a cladistic analysis that knocked *Archaeopteryx* from its special perch on the bird lineage, relegating it to a different branch along with a host of other feathered dinosaurs. That study has met resistance from some other palaeontologists, who question the strength of the cladistic analysis and say that the evolutionary relationships will remain unclear until more early birds and their close relatives are discovered.

The Liaoning fossils have led Xu to make other bold proposals about the origins of flight. The discoveries of *Microraptor* and *Anchiornis*, another four-winged dinosaur, led Xu to argue[6] that the four-winged trait was not an evolutionary dead end, as had been previously assumed, but could actually have been the transitional step between dinosaurs and birds.

The feathered dinosaur fossils have also provided some of the first hard evidence for when

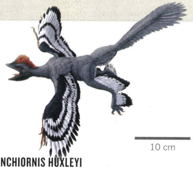
## XING XU'S FEATHERED FRIENDS

Fossils found in Liaoning in northwestern China show that many dinosaurs in the late Jurassic and early Cretaceous periods had feathers. The exceptional specimens have transformed ideas about theropod dinosaurs and the birds that evolved from them.



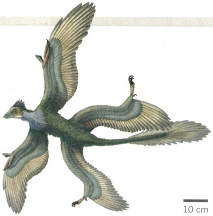
**XIAOTINGIA ZHENGI**

*Late Jurassic (160 million to 145 million years ago)*
The 30-centimetre-long *Xiaotingia* had feathers and other features resembling those of *Archaeopteryx*, often considered the earliest bird. Xu has proposed that both belonged to a group of non-avian dinosaurs, closely related to but distinct from birds.



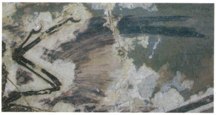
**ANCHIORNIS HUXLEYI**

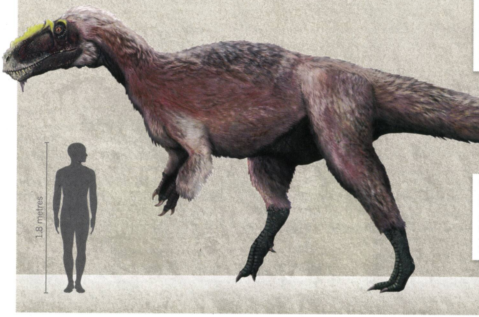
*Late Jurassic (160 million to 145 million years ago)*
An exquisitely preserved specimen of the dinosaur *Anchiornis* helped Xu and his colleagues pin down the timing of the transition from dinosaurs to birds. Its long feathers demonstrated how complex early plumage could be.



**MICRORAPTOR GUI**

*Early Cretaceous (145 million to 100 million years ago)*
Although not a bird, the tiny dinosaur *Microraptor* had feathers on its arms (see below) and on its legs, and it may have flown.

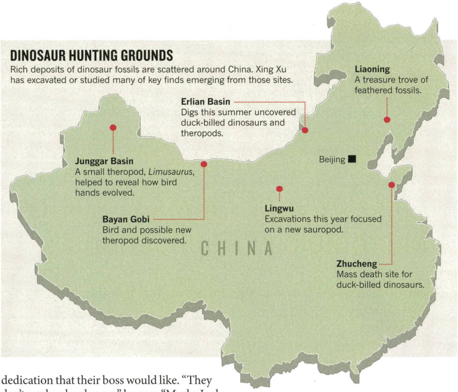




**YUTYRANNUS HUALI**

*Early Cretaceous (145 million to 100 million years ago)*
This 9-metre-long long predator provided evidence that even some big dinosaurs had feathers. Three specimens were found — two juveniles and an adult — with feathers in various locations, including the hip, neck and back.

*X. ZHENGI*: XING LIDA/LIU YI; *A. HUXLEYI*: J. T. CSOTONYI/SPL; *M. GUI* RECONSTRUCTION: P. SLOAN; *M. GUI* PHOTOGRAPH: REF. 3; *Y. HUALI*: BRIAN CHOO



**DINOSAUR HUNTING GROUNDS**

Rich deposits of dinosaur fossils are scattered around China. Xing Xu has excavated or studied many of key finds emerging from those sites.

Liaoning
A treasure trove of feathered fossils.
Erlian Basin
Digs this summer uncovered duck-billed dinosaurs and theropods.
Junggar Basin
A small theropod, Limusaurus, helped to reveal how bird hands evolved.
Beijing
Bayan Gobi
Bird and possible new theropod discovered.
Lingwu
Excavations this year focused on a new sauropod.
CHINA
Zhucheng
Mass death site for duck-billed dinosaurs.

and why feathers evolved. “For most of the past century, the classic issue in feather evolution was that the fossil record told us essentially nothing,” says Richard Prum, who studies the evolution of birds at Yale University in New Haven, Connecticut. “What's happened with the Liaoning formation has been a totally new chapter.”

In the past, palaeontologists had presumed that when feathers first arose, they helped bird ancestors to fly. But on the basis of his discoveries, Xu makes the controversial argument that most dinosaurs probably had at least a smattering of plumage, which would mean that feathers originally served other functions, such as attracting mates or insulating against the cold.

## ROLE MODEL

With his extraordinary track record, Xu brings to mind the prodigious US palaeontologists Othniel Marsh and Edward Cope, who discovered dozens of dinosaurs in the late nineteenth century in a frenetic competition that became known as the bone wars. But whereas those Victorian fossil hunters made frequent errors, such as giving new names to species that had already been described, Xu is a careful researcher who does not rush into print. His published record of new species has rarely been challenged, says Mike Benton, a palaeontologist at the University of Bristol, UK, who has analysed the accuracy of dinosaur researchers.

Xu would like to see Chinese science as a whole become more careful. “Chinese culture is a problem for science because it's not logical enough,” says Xu during a trip this summer, as his driver gaily overtakes on the wrong side of the road. “Traditionally people don't like to criticize, either. For peer review you have to criticize in some way...”. He breaks off mid-sentence to answer a phone call in Mandarin for a few minutes, before resuming exactly where he left off “... but here in China we don't have a real peer-review system.”

Another problem hanging over Chinese palaeontology is fakery. Xu is keenly aware of it. In 2000, he helped[7] to unmask one of the biggest hoaxes in a generation: a composite specimen named *Archaeoraptor*, made up of the upper body of an ancient bird and the tail of the dinosaur *Microraptor*. Scientists are getting better at spotting fakes, says Xu, but they do still crop up, because poor farmers know that they can sell the most unusual fossils to museums or institutes for hefty sums. “We have the greatest resources in palaeontology now,” says Zhonghe Zhou, director of the IVPP, “but on the other hand, the destruction of localities, the faking — those kinds of things are often the most severe. The law isn't good enough.”

Xu worries about the future of his profession, particularly the next generation of scientists. His current students aren't showing the dedication that their boss would like. “They don't work as hard as me,” he says. “Maybe I ask too much, maybe that's my problem.” Qing-Jin Meng, director of the Beijing Museum of Natural History, says, “Excellent palaeontologists [such as Xu] are hard to find.” Part of the problem may be the globalization of Chinese palaeontology, he adds. “Many students who have great potential have gone to the United States and European countries to study.”

Xu says that if only he could find the time, he would like to write articles about how to improve Chinese science. But so far he has published only one blog post in Mandarin.

**HE IS THE GO-TO MAN IN CHINA FOR ANYTHING PEOPLE WANT TO KNOW ABOUT DINOSAURS.**

“Honestly, I don't like it much. I'd rather do science,” he says.

Xu's packed schedule can be hard on his family — his wife Zhonghia Zhou, who is a secretary at the Institute for Geology and Geophysics in Beijing, and their two boys, aged 7 and 12. “My wife complains because the kids are growing up,” confesses Xu. “She says they need a male example. And I thought, yeah, that's important.”

So in the past couple of years he has tried to spend more time at home, helping with homework, playing table tennis with his wife and taking his family on days out to Beijing's parks. Even the director of the IVPP recognizes a candidate for burn-out when he sees one: “He should slow down a bit!” says Zhou. “You can't study everything — you need time for hobbies.” To that end, Xu and Zhou sometimes play badminton on the court installed in the entrance hall of the IVPP.

It is unlikely that Xu's hobbies will eat into his prodigious output too much. Back in his office in Beijing after the trip to Zhucheng, Xu rummages through the floor-to-ceiling cupboards lining two walls. He pulls out slabs of rock, pointing out salient features and clues that he might have an unknown species on his hands.

More than setting records by finding new creatures, Xu is interested in asking and answering questions about a far-gone era, when his country was filled with a dizzying array of feathered dinosaurs and birds. He is keen, for example, to continue exploring how non-avian dinosaurs developed feathers and whether the plumage differed from that of modern birds. As he looks over the fossils in his office, Xu's eyes glint with a blend of tiredness and excitement.

Luo, who has watched Xu's career take off, sees no end to the potential discoveries. “Fossils are silent,” says Luo. “It takes an insightful palaeontologist to tell their story, and Xu Xing is a fantastic storyteller.” ■

---

**Kerri Smith** *is podcast editor for* Nature *in London.*

1. Xu, X., You, H., Du, K. & Han, F. *Nature* **475,** 465–470 (2011).
2. Xu, X. *et al. Nature* **484,** 92–95 (2012).
3. Xu, X. *et al. Nature* **421,** 335–340 (2003).
4. Xu, X. & Norell, M. A. *Nature* **431,** 838–841 (2004).
5. Zhao, X., Cheng, Z. & Xu., X. *J. Vert. Paleontol.* **19,** 681–691 (1999).
6. Xu, X. *et al. Chinese Sci. Bull.* **54,** 430–435 (2009).
7. Xu., X., Zhou, Z. & Wang, X. *Nature* **408,** 705–708 (2000).

# nature.com webcasts

**MACMILLAN**
**SCIENCE COMMUNICATION**

*Exclusive partner of Nature Publishing Group, publisher of* Nature and Scientific American

presents a custom webcast on:

# The Power of Flow Cytometry in Microbiology Applications

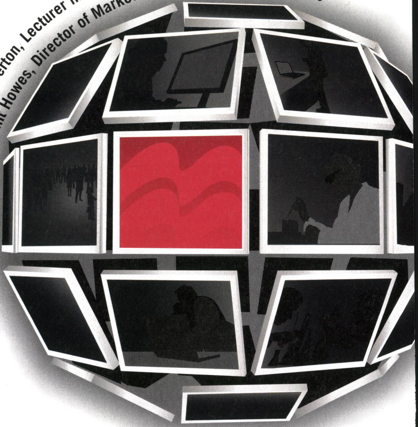
Please join us for a lively discussion on the benefits of using flow cytometry in diverse microbiological applications.

**Date:**
**September 27th 2012**

**Time:**
**1pm Eastern / 10 a.m. Pacific**
**6p.m. GMT/ 7.p.m CET**

*available on demand from September 28th*

**Speakers: Dr Tim Overton, Lecturer in Biochemical Engineering**
**Grant Howes, Director of Marketing, Personal Flow Cytometry**



Register for **FREE:**
**www.nature.com/webcasts**

Sponsored by: BD

*BD Biosciences is fully responsible for all editorial content*

nature publishing group npg

# COMMENT

**ARCHITECTURE** The Victorian woman who celebrated nature in stone **p.29**

**MUSEUMS** Mathematician Glen Whitney talks about founding MoMath **p.32**

**SANITATION** A call to stop India's trains spreading disease across the subcontinent **p.33**

**OBITUARY** Martin Fleischmann, electrochemist in cold fusion furore, remembered **p.34**

B. PIERCE/TIME LIFE PICTURES/GETTY

**The town of Times Beach in Missouri was evacuated in 1983 and later demolished after a dioxin spill.**

# Rethink chemical risk assessments

## The US Environmental Protection Agency needs to speed up its risk analyses and address uncertainty, say **George M. Gray** and **Joshua T. Cohen**.

The US Environmental Protection Agency (EPA) is under fire. Its flagship Integrated Risk Information System (IRIS), which develops risk values for human chemical exposure that are used by regulators and others, is being widely criticized for being too slow and scientifically flawed. The system needs an overhaul.

Last year, for instance, the US National Academy of Sciences (NAS) castigated the EPA's inadequate assessment of the health risks of formaldehyde[1]. Evaluations of other chemicals, including dioxin, have been equally controversial[2]. In December 2011, Congress directed the agency to improve its risk assessments and submit documentation to the NAS for review (see go.nature.com/xmeqyv). But the problems go deeper than the IRIS process.

Two main challenges render the EPA's risk assessments inadequate for decision-making. First, they take years or even decades to conclude, meaning that many chemicals have never been examined. Second, their scientific credibility is often challenged. Peer reviewers have questioned the EPA's selective use of data and some assumptions that it has made to plug gaps in the scientific evidence. The NAS has recommended that the EPA better justify and quantify its risk-assessment assumptions.

As scientists who have served at the EPA (G.M.G.) and participated in NAS reviews (J.T.C.), we believe that more is needed. The agency needs to fundamentally alter its approach to risk evaluation. First, it should offer faster summaries for more chemicals. Rough-and-ready estimates are often sufficient for policy-making, and are better than nothing. IRIS should include information from private groups and other governments, and apply available techniques for calculating the risks of chemicals for which there are few data. Second, the EPA needs to acknowledge that its risk estimates are uncertain by reporting a range of plausible values, not just those that support its science-policy goals.

### ROOTED IN THE PAST

Attitudes towards environmental regulation have changed since the agency was founded in 1970. Less than a decade after Rachel Carson exposed the environmental damage caused by the pesticide DDT in her 1962 book *Silent Spring*, Americans wanting "freedom from risk"[3] embraced government protection.

The EPA successfully addressed health threats posed by high-profile pollutants. A ban on leaded petrol spearheaded by the EPA in 1973 helped to reduce the level of lead in children's blood by nearly an order of magnitude in the decades that followed. Other agency regulations introduced in the early 1970s halved the levels of air pollutants such as sulphur dioxide and carbon monoxide.

By the mid-1990s, the most glaring environmental problems had been dispatched and the EPA's progress stalled. Although IRIS now counts 557 finished risk assessments in its repository, releases in each year since 1995 have mostly been in single digits ▶

▶ (see 'Count down'). Risk assessments have become mired in controversy and extended review cycles. Worse, the EPA prioritizes revisions to assessments of chemicals it has already evaluated, such as dioxin and mercury[4], rather than evaluating crucial chemicals for the first time.

The slow pace of IRIS threatens public health. Many people might assume that chemicals lacking an IRIS risk estimate are safer than those that have been assigned one, even if they are not. For example, the EPA's assessment of perchloroethylene, used in dry cleaning, has encouraged phasing out of the chemical. Some dry cleaners are switching to *n*-propyl bromide — for which there is no IRIS entry — despite evidence that it may pose a greater health risk than perchloroethylene[5].

Other difficulties arise from EPA efforts to characterize risk at ever-lower exposure levels, at which health effects are hard to observe. Reliant on animal experiments, the agency resorts to two critical assumptions: that any adverse health effects seen in rodents are mirrored in humans, and that the high doses used in the lab (to see an effect using a reasonable number of animals) can be extrapolated downwards, often by orders of magnitude, to reflect human population exposures. As the NAS has pointed out, the EPA often fails to justify the data used or explain how risks were estimated at low levels[1,2].

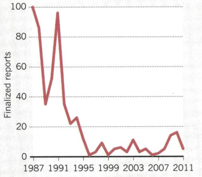
In our view, the problem is the EPA's use of assumptions that it claims are "public health protective", which err on the side of overstating risk when data are lacking. Take dioxin, for example. In its assessment, the EPA assumed the worst case — that low levels of dioxin cause cancer — because that possibility cannot be ruled out. Yet other agencies, including the World Health Organization[6], interpret the biological studies of dioxin as suggesting that it is unlikely to cause cancer at low levels because of the way the chemical behaves within cells.

Such inflated risk estimates can lead to overly stringent regulations and can scramble agency priorities because the degree of precaution differs across chemicals. For example, the EPA's National-Scale Air Toxics Assessment from 2005 estimated a tenfold-higher cancer risk from outdoor air exposure to carbon tetrachloride (used in dry cleaning and as a solvent and refrigerant) than from ethylene dibromide (a termite fumigant and former additive in petrol). Yet by taking on board the biological evidence, other agencies around the world have concluded the opposite — that carbon tetrachloride poses little risk because, unlike ethylene dibromide, it has a threshold for its carcinogenic action.

The EPA intended that its air-toxicity results would help to set priorities for improving data in emission inventories, to target risk-reduction activities more effectively and to identify pollutants and industrial sources of greatest concern. But its aggressive use of precautionary assumptions, even when they are scientifically unwarranted, instead misleads decision-makers.

**COUNT DOWN**

Chemical risk assessments completed by the US Environmental Protection Agency have stalled since the mid-1990s.



SOURCE: EPA

### THE WAY FORWARD

To its credit, the EPA has committed to adopting the NAS recommendations, including streamlining presentation of its analyses, making its toxicity evaluations more uniform and incorporating multiple data sets[7]. To become fit for purpose again, the agency must change its view of risk assessment. It should not see assessments as a search for scientific truth, but as a way to bring available information to bear on regulatory and public-health decisions.

The EPA should expand IRIS to include sources of information that are not currently used, similar to the International Toxicity Estimates for Risk Assessment database (www.tera.org/iter). IRIS should report risk values developed by international public-health agencies, by other health agencies in the United States and by private groups.

The agency should integrate into IRIS information from its internal programmes, such as its Provisional Peer-Reviewed Toxicity Value database, which contains more than 300 rapid-risk estimates developed to inform clean-up decisions at hazardous-waste sites. These estimates draw on information of varying quality, such as short-term toxicity tests, expert judgements and statistical models that predict a chemical's behaviour on the basis of its structure. The associated uncertainties should be reflected in the IRIS entry.

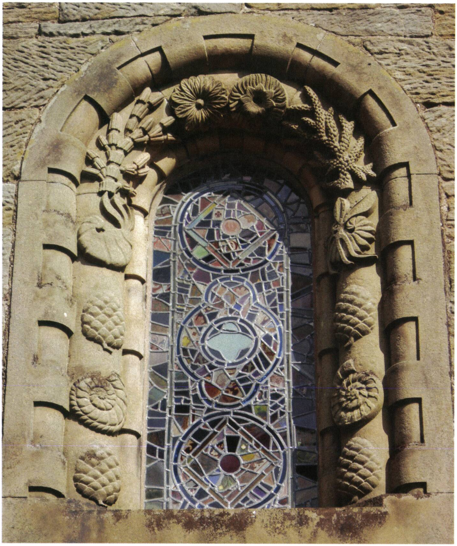
In the longer term, the EPA should expedite its ongoing exploration of high-throughput screening methods. These can quickly ascertain a broad range of properties for a chemical, such as how readily it reacts with biological systems, and hence evaluate potential health risks[8]. Once these methods and an understanding of how they feed into risk estimates are established, the information should be incorporated into IRIS.

Fundamentally, the EPA should replace risk values that are built on science-policy assumptions with risk estimates that acknowledge underlying uncertainties. For instance, the agency could follow the example of the Intergovernmental Panel on Climate Change[9] and report a range of risks that correspond to different models. Users would then be able to see whether a value is sufficiently precise to support a particular course of action.

Critics might argue that decision-makers will suffer 'paralysis by analysis' if confronted with a range of values rather than just one. Yet that is how it should be. The EPA's definitive values are illusions: they conceal uncertainty that cannot be resolved scientifically. Bringing conflicting value judgements into the open will enable honest debate and improve public health. ■

**George M. Gray** *is director of the Center for Risk Science and Public Health at George Washington University, Washington DC 20037, USA. In 2005–09, he was assistant administrator to the EPA Office of Research and Development and EPA science adviser.*
*e-mail: gmgray@gwu.edu*
**Joshua T. Cohen** *is deputy director of the Center for the Evaluation of Value and Risk in Health at Tufts Medical Center, Boston, Massachusetts 02111, USA. He served on the National Academies committees that reviewed the EPA's dioxin risk assessment in 2006 and its risk-assessment methodologies in 2009.*
*e-mail: jcohen@tuftsmedicalcenter.org*

1. National Research Council. *Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde* (National Academies Press, 2011).
2. National Research Council. *Health Risks from Dioxin and Related Compounds: Evaluation of the EPA Reassessment* (National Academies Press, 2006).
3. Sunstein, C. R. *Risk and Reason: Safety, Law, and the Environment* (Cambridge Univ. Press. 2002).
4. US Environmental Protection Agency. *US Federal Register* **77**, 26751–26755 (2012).
5. Finkel, A. M. *Increased Toxicity and Carcinogenicity of n-Propyl Bromide (1-Bromopropane) Relative to Perchloroethylene.* Supplemental Report to The City Of Philadelphia Department of Public Health/Air Management Services (2010); available at http://go.nature.com/nr3tev.
6. World Health Organization. *Fact Sheet No. 225: Dioxins and their Effects on Human Health* (WHO, 2010); available at http://go.nature.com/m3deqr.
7. US Environmental Protection Agency. *EPA's Integrated Risk Information System Program: Progress Report and Report to Congress* (EPA, 2012); available at http://go.nature.com/62fvu4.
8. Collins, F. C., Gray, G. M. & Bucher, J. R. *Science* **319**, 906–907 (2008).
9. Carter, T. R. *et al.* in *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Parry, M. L. *et al.*) 133–171 (Cambridge Univ. Press, 2007).



RAYMOND WHITTAKER

**Pinecones, flowers and ammonites adorn the windows of Saint Mary's church in Wreay, UK.**

ARCHITECTURE

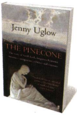
# Life in stone

**Georgina Ferry** enjoys a biography of a little-known Victorian woman who built monuments to nature.

In the village of Wreay, just south of the border between England and Scotland, stands a wholly original building: a synthesis of decorative motifs drawn from early nineteenth-century geology and natural history with an ancient architectural style. This small church, completed in 1842, is the work of a remarkable Victorian, Sarah Losh.

As Jenny Uglow reveals in her intriguing biography, *The Pinecone*, Losh was by the age of 18 a competent mathematician, linguist and classicist, and knowledgeable about science, architecture, politics, philosophy, literature and art. Her nineteenth-century biographer, Henry Lonsdale, wrote: "With powers to grapple with Euclid and algebra, she had but to give her attention to any subject to master it." She also had a clear sense of her own self-worth. Unlike writers of her time such as Jane Austen, Mary Shelley and Elizabeth Gaskell, she has not achieved worldwide recognition. Yet after her death, Dante Gabriel Rossetti hailed her as a genius, and her work foreshadowed the designs of John Ruskin, Alfred Waterhouse and William Morris.

Books move, but buildings stay in one place. Losh, by building almost exclusively in Wreay, ensured that beyond her immediate locality only specialists would come to know and admire her work. Panning outwards from this small, largely agricultural community, Uglow uses Losh's story to create a vibrant panorama of early nineteenth-century society that extends throughout the British Isles, across Europe and even to the deadly passes of Afghanistan. Uglow is at ease in the intellectual environment of the era, which she researched fully for her book *The Lunar Men* (Faber, 2002).



**The Pinecone: The Story of Sarah Losh, Forgotten Romantic Heroine — Antiquarian, Architect and Visionary**
JENNY UGLOW
*Faber/Farrer, Strauss and Giroux: 2012/2013.*
*344 pp./352 pp.*
*£20/$28*

Losh's family of country landowners provided wealth, stability and an education infused with principles of the Enlightenment. Her father, John, and several uncles were experimenters, industrialists, religious nonconformists, political reformers and enthusiastic supporters of scientific, literary, historical and artistic endeavour, like members of the Lunar Society in Birmingham, UK. John Losh was a knowledgeable collector of Cumbrian fossils and minerals. His family, meanwhile, eagerly consumed the works of geologists James Hutton, Charles Lyell and William Buckland, which revealed ancient worlds teeming with strange life forms.

Sarah's uncle James Losh — a friend of political philosopher William Godwin, husband of the pioneering feminist Mary Wollstonecraft — took the education of his clever niece seriously. She read all the latest books, and met some of the foremost innovators of the day, such as the mathematician Isaac Milner and the physicist John Leslie.

On their father's death in 1814, Sarah and her beloved sister Katharine inherited substantial property in Wreay and interests in their father's successful alkali factory in the expanding industrial city of Newcastle. Their financial independence secure, neither ever married. Instead, they toured France, Germany and Italy together. In Italy, Losh saw for herself the simplicity of classical Roman and early medieval architecture. Once home, the sisters built a school and a house for the local schoolmaster based on simple, pre-Renaissance forms — the house was a copy of a Pompeiian cottage. After Katharine died, Losh embarked on her masterpiece.

Brooking no argument from the Bishop of Carlisle, she offered to fund the complete rebuilding of Saint Mary's, her village church, on the condition that she "be left unrestricted as to the mode of building it". ▶

She ignored the contemporary craze for the Gothic, opting instead for a style modelled on the Romanesque: a simple rectangular building with a semicircular apse, and doors and windows topped with round arches.

She made the building entirely her own by adding decorative carvings that combined rich pre-Christian symbolism with natural forms recently brought to light by fossil-hunters and naturalists. Executed by local craftsmen (and sometimes Losh herself) working mostly in local stone and wood, these anticipated the artistic and architectural ideals set out by John Ruskin a decade after the church was completed. Lotus flowers, ammonites and butterflies embellished windows, doorways and capitals; Losh filled the high windows of the apse with the delicate forms of local fossil ferns cut from translucent sheets of alabaster. More than 30 years after she completed her church, and on a much grander scale, Alfred Waterhouse adopted a Romanesque design decorated with flora and fauna for the Natural History Museum in London. Like Losh, he was inspired by visiting Italy and studying natural history, but Uglow cites no evidence that he knew of Losh's work.

Losh's carvings often feature a pinecone, an ancient symbol of regeneration and enlightenment. Uglow points out that the number of spirals winding up from the base of a pinecone always belongs to the Fibonacci series (running 1, 2, 3, 5, 8 and so on, without end). James Hutton memorably concluded that he could find "no vestige of a beginning, no prospect of an end" in his studies of geological strata. Uglow helps us to see how Losh combined the architectural evidence of past human societies with contemporary invention and discovery, and how she conveyed, through her buildings, a sense of the eternal.

Most of Losh's personal papers and journals, like those of Jane Austen, were lost or destroyed, leaving the biographer to piece together her life from fragments gleaned elsewhere. Sarah Losh remains something of an enigma: a deeply religious woman who built a church that contained no overtly Christian symbols; a devotee of ancient structures and a daughter of the Industrial Revolution; a fashionable beauty and an unmarried scholar and craftswoman.

Sarah Losh chose to express herself in stone, rather than words. In Jenny Uglow, she has found a fine interpreter. ■

**Georgina Ferry** *is a science writer and author living in Oxford, UK.*
*e-mail: mgf@georginaferry.com*

---

Canadian pianist Glenn Gould recorded Bach's *Goldberg Variations* twice, in 1955 and 1981.

G. PARKS//TIME LIFE PICTURES/GETTY

TECHNOLOGY

# Baroque geekery

**Tim Boon** assesses a take on the evolving technology behind recordings of J. S. Bach.

**Reinventing Bach**
PAUL ELIE
*Farrar, Straus and Giroux: 2012*
*496 pp. £19.99, $30*

Paul Elie reveres the music of J. S. Bach and loves some recordings in particular, such as Glenn Gould's 1955 rendition of the *Goldberg Variations*. In *Reinventing Bach* Elie sets out to show how technologies — especially developments in recording — have been central to the twentieth century's experience of "the Master's" music.

The book's conceit is that the composer of the *Two- and Three-Part Inventions* was in some sense an inventor, and so peculiarly attuned to being reinvented — through the recording technologies of the past 100 years or so. And, as Elie shows, the power that recording offered, of enabling repeated listening, also accelerated the rediscovery of Bach by generations of musicians.

Each chapter takes a key recording, dwelling to different degrees on the technology used — disc, tape or digital. The chapters are arranged in roughly chronological order and range from takes by Albert Schweitzer and Leopold Stokowski on the famous *Toccata and Fugue in D Minor* to Gould's two recordings of the *Goldberg Variations* and beyond. Alongside this, Elie threads a biography of Bach, period-setting snapshots of cultural events and an accumulating cast of Bach performers and recording artistes.

Throughout, Elie describes the music, not with the technical terminology of the conservatoire, but with metaphor and simile. His characterization of the *Toccata and Fugue in D Minor*, for instance, reads: "the pipes ring out once, twice, a third time. Then with a long, low swallow the organ fills with sound, which spreads toward the ends of the instrument and settles, pooling there." What he doesn't do, however, is meet the promise in the publisher's blurb to give us "a nuanced and intelligent examination of the technology" that has made the reinvention of Bach possible.

➲ NATURE.COM
For more on recording music, see:
go.nature.com/xx3x22

Elie draws on a wide range of published literature, and is insightful about the interplay between technological change and the development of both individual technique and the market for classical music. For example, he describes how Gould's recordings of the *Goldberg Variations* were polished as the pianist, holed up at a country retreat, repeatedly recorded and listened back to his own performances of the 30 variations on the recently invented tape recorder. Elie also nicely depicts how the historically informed performance scene was stimulated by the arrival of the CD: the clarity of digital recording gave period-music specialists an opportunity to provide newly 'authentic' performances.

But the descriptions of technologies are less sure. Magnetic recording tape does not use silver oxides, as the book has it, but iron oxides. Elie also writes that Schweitzer recorded on cylinders, yet EMI always used discs. His description of a 1905 Victrola gramophone as having a needle converting movements to electrical impulses reads oddly. This is an entirely acoustic device in which even the motor is clockwork; there were no electrical gramophones before the 1920s.

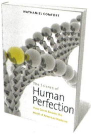
The book would also be stronger for a deeper and more integrated account of musical instruments. The hybrid instrument given to Schweitzer by the Paris Bach Society when he went as a missionary to Africa — enabling him to play in tropical conditions — is described merely as having "the features of a piano and an organ: two manuals, strings and hammers, pedals. The inside of it was lined with zinc to ward off moisture in the tropics". (This amazing-sounding machine can be seen in the Maison Albert Schweitzer, the organist's former home, in Alsace, France.) Similarly, Bach's possible involvement in the development of a new instrument called the Lautenwerck, a kind of keyboard-actuated lute, is glossed over in two brief paragraphs — a loss, given the emphasis on Bach as inventor.

> ***"The pipes ring out once, twice, a third time. Then, with a long, low swallow the organ fills with sound."***

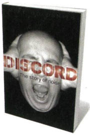
In the end, *Reinventing Bach* reads best as a sincere and compelling account of the author's love of Bach's recorded oeuvre. The passion shines through even though the technology is more marginal than promised. And you may find yourself compelled to rummage through your CD shelves for the works — as I did — revisiting Bach in his multifarious reinventions. ■

**Tim Boon** *is head of research and public history at the Science Museum in London, UK.*
*e-mail: tim.boon@sciencemuseum.ac.uk*

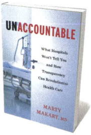
# Books in brief



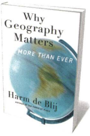
**The Science of Human Perfection: How Genes Became the Heart of American Medicine**
*Nathaniel Comfort* YALE UNIVERSITY PRESS *336 pp. £25 (2012)*
In this provocative look at genetic medicine in the United States, medical historian Nathaniel Comfort argues that eugenics casts a long shadow over the field. He has researched records spanning a century, following the ever-evolving group of geneticists, eugenicists, psychologists, medics, public-health workers, zoologists and statisticians intent on using heredity to improve human life. Today's hybridized discipline, he says, is noble in intent but rife with social and ethical questions centred on the 'illusion of perfectibility'.



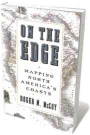
**Discord: The Story of Noise**
*Mike Goldsmith* OXFORD UNIVERSITY PRESS *336 pp. £16.99 (2012)*
You might pay to hear a jazz saxophonist let rip in a club, but go crazy if they practised next door. Sound in the wrong place is noise, points out science writer and former head of acoustics at the UK National Physical Laboratory Mike Goldsmith in this chronicle of cacophony and our attempts to control it. Starting with the nature of sound and its birth in the infant Universe, he runs through prehistoric noise, the beginnings of acoustical science in the Renaissance, the machine-led din of the Industrial Revolution, the clamorous twentieth century and today's aural pollution from wind farms, underwater sonar and more.


UNACCOUNTABLE
What Hospitals Won't Tell You and How Transparency Can Revolutionize Health Care
MARTY MAKARY, MD

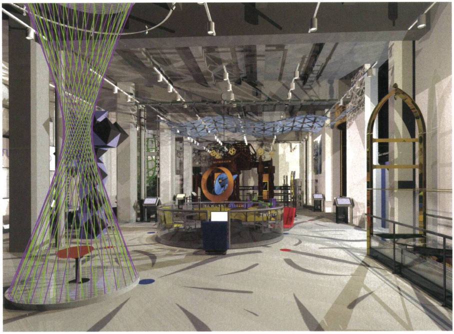
**Unaccountable: What Hospitals Won't Tell You and How Transparency Can Revolutionize Health Care**
*Martin Makary* BLOOMSBURY *256 pp. £19.99 (2012)*
Surgeon and health-policy specialist Martin Makary reveals US hospitals as battlegrounds between competence and chaos. Serious blunders — such as surgical tools being left in body cavities — are so common that a 2010 study reported that one-quarter of patients are harmed by medical mistakes. Among Makary's mind-bending observations is how two doctors approached the removal of benign colonic polyps. One neatly excised the growth; the other removed half the colon. A powerful plea for openness in US health care.



**Why Geography Matters, More Than Ever**
*Harm de Blij* OXFORD UNIVERSITY PRESS *320 pp. £10.99 (2012)*
Where geopolitics is concerned, Harm de Blij says, it's easy to hit a *plus ça change* moment. This revised edition of his influential 2007 book includes the rapid shifts and upheavals of the past five years, from the Arab Spring to the European Union's economic wobbles. But de Blij's original premise — that the geographical illiteracy prevalent in the United States seriously impedes coherent policy — is more relevant than ever. With power comes responsibility, and Americans, he says, have an obligation to develop the geographer's perspective on culture, politics, economics and the environment.



**On the Edge: Mapping North America's Coasts**
*Roger M. McCoy* OXFORD UNIVERSITY PRESS *256 pp. £18.99 (2012)*
Some 500 years ago, the edges of North America were as mysterious to Europe's explorers as the Moon. Geographer Roger McCoy recounts their voyages and cartographic efforts, starting with John Cabot and Martin Frobisher, and ending with Otto Sverdrup and Vilhjalmur Stefansson in the early twentieth century. The tales of derring-do, brushes with death and brutal behaviour towards native Americans are interspersed with clear explanations of how, over time, this multitude of mariners redrew the New World map.



A rendering of the upper level of the Museum of Mathematics, which opens in New York later this year.

MUSEUM OF MATHEMATICS

## Q&A Glen Whitney

# Maths demystifier

*Mathematician Glen Whitney left a job in finance to set up the Museum of Mathematics (MoMath), which is due to open in Manhattan, New York, on 15 December. He wants to spread the word that mathematics is a beautiful discipline and all around us, from the geometry of soap bubbles to the algorithms that control traffic lights.*

N. HIGGINS

**How did you start out in mathematics?**

When I was young, I broke my collarbone playing soccer and fell in love with maths problems while recuperating. I had a voracious appetite for mathematics in high school but when I went on to Harvard had no illusion that I was going to be one of the top researchers in the country. After teaching at the University of Michigan, I received an offer to try statistical trading at Renaissance Technologies in New York, a hedge fund run by mathematician Jim Simons. I decided to give it a try. I started out in the data group, then migrated into researching trade strategies and on to improving the research tools themselves. It was exciting and intellectually demanding, but I wanted to do something beneficial to society at large.

**Why did you focus on the public image of mathematics?**

The National Security Agency views the shortage of US mathematicians as one of the country's biggest security threats. Yet you often hear people say, "I was always terrible at maths". No one says that about reading. I believe this attitude stems primarily from the emphasis on rote procedures and people paying too little attention to making connections with everyday life and the world around them. We need a cultural institution to combat this prejudice.

**And why a museum?**

Many science museums are sparse on maths content. A lack of contemporary mathematics exhibits means that one from 1960 is still housed at the New York Hall of Science and the Museum of Science in Boston, Massachusetts. When kids see chemistry and physics exhibits but none on mathematics, it conveys a subtle but powerful message. The United States used to have one museum of mathematics — on Long Island, New York — but it was so small you had to gather ten people for it to open. It closed down in 2006 and I realized that was an opportunity to create an environment for people to have seminal experiences with mathematical concepts, to show that maths is as much a part of our society as the other sciences.

**What will a visitor find at your museum?**

Hands-on exhibits showing how mathematics can be tangible, open-ended and fun. In the new museum, we will have exhibits on everything from the beautiful patterns created by video feedback to the probabilities of making a free throw in basketball.

**Have you debuted any of your exhibits?**

Yes, in a travelling exhibition, the *Math Midway*, which has appeared at science museums and festivals across the country, and will continue to tour beyond the museum opening. Its iconic exhibit is a square-wheeled tricycle that rides smoothly over a surface of inverted catenary curves calculated to keep the axles of the tricycle level as the corners of the wheel rotate, which seems to give people the sense that maths can make the impossible possible. Another exhibit is a plane of laser light that shows all of the possible cross-sections of translucent three-dimensional solids. Visitors can rotate a cube to learn that it can be sliced to yield not just squares and triangles, but trapezoids, rhombuses and even a regular hexagon that cuts through all six faces.

**What else are you doing until the museum opens?**

MoMath holds a monthly lecture series called Math Encounters in which we strive to show unexpected ways that maths touches everyday life — such as in the geometry of soap bubbles. Upcoming presentations include a talk about the maths of sport, and one on the maths of origami.

**What sets MoMath apart from other mathematics outreach and education efforts?**

Besides the fact that MoMath will be the only museum in North America devoted specifically to mathematics, there are a few distinctive aspects to its approach: a focus on physical interaction, especially whole-body involvement; an effort to show as broad a spectrum of the world of mathematics as possible, not tied to any specific curriculum; and an emphasis on giving people the experience of the "Aha!" moment of discovery.

**You also run mathematical walking tours in Manhattan. What do those involve?**

I talk about the algorithms used to control traffic lights, the mathematical issues involved in keeping the subway running, the symmetry of the mouldings on the sides of buildings and the unusual geometry that gives gingko trees their distinctive shape. There are deep connections to music, art and finance. If you give me a route, I'll make a tour. There is maths everywhere.

**INTERVIEW BY JASCHA HOFFMAN**

# Correspondence

## NASA bids are not a popularity contest

You recently conducted an online popularity poll of three proposals competing for selection as the next NASA Discovery Program mission (*Nature* http://doi.org/h79; 2012). In my view, the concept and execution of this poll demeans *Nature* and belittles what is at stake.

Worse, there are indications that the poll could have been manipulated. Voting for one particular mission occurred in a large burst on a single day. It is immaterial whether this was caused by the mission teams enlisting many supporters to vote quickly, or by people who worked out an easy way to vote multiple times. The point is that the results are not meaningful.

Popularity contests are not the way to choose among scientific alternatives. Although public interest needs to be taken into account when spending taxpayers' money, selecting a mission should ultimately depend on its scientific merit and technical feasibility.

NASA's missions have a track record of exciting the public anyway, with web hits for different missions leading to server saturation during key events. The likely effectiveness of each mission's outreach programme needs to be evaluated by looking carefully at the large, detailed proposals submitted by each mission team.

**Michael F. A'Hearn** *University Park, Maryland, USA.*
*mahearn@mac.com*

*Competing interests declared; see go.nature.com/1qferq.*

## Tourism ban won't help Indian tigers

The Indian Supreme Court's temporary injunction against tourism in core areas of tiger reserves could place the animals at greater risk of poaching if it becomes permanent, by reducing revenue for park management (*Nature* **488,** 10; 2012). The injunction has now been extended until 27 September.

Most of the reserves with the highest numbers of tigers and tourists are in the state of Madhya Pradesh. In 2010–11, the state's 35 parks received US$17.1 million from government sources. Five tiger reserves generated most of the $2.8 million obtained from tourism. In 2011–12, Bandhavgarh reserve received $1.2 million in tourist revenue and almost the same amount from government sources. Tourism therefore yields 25–50% of tiger conservation funds in Madhya Pradesh, safeguarding up to 130 tigers.

Different management strategies would be more effective in overcoming conservation concerns stemming from disruptive tourist behaviour.

**Ralf C. Buckley** *Griffith University, Australia.*
*r.buckley@griffith.edu.au*
**H. S. Pabla** *Madhya Pradesh, India.*

## Tighten up Japan's stem-cell practices

Japan has bioethical regulations and clinical guidelines in place for experimental stem-cell therapies and for stem-cell-based pharmaceuticals. As a forensic pathologist who has worked on a patient who died after mesenchymal stem-cell therapy in Japan, I am aware that other patients receiving this treatment have developed serious and even fatal complications. These cases indicate that Japan's regulatory infrastructure needs to be more strongly enforced.

Reaction in Japan to these cases has been minimal. This contrasts with the tough approach of the US Food and Drug Administration, which led to the prompt prosecution of clinicians and companies involved in similar cases in Colorado and Texas (see *Nature* **477,** 377–378; 2011).

Japan's Investigative Commission for Institutional Framework in Regenerative Medicine recommended establishing a punitive system for physicians and clinics practising unethical activities, but its 2011 report made no mention of such plans. The country's specialist medical organizations should push for government collaboration if an effective disciplinary system is to be established (E. Dolgin *Nature Med.* **16,** 495; 2010).

The Japanese Medical Ethics Committee, for example, needs to work more like the UK General Medical Council, which does not depend on the country's judiciary system to exercise its powers.

The Japanese Society for Regenerative Medicine and the International Society for Stem Cell Research should collaborate with Japan's health ministry to establish a system to prevent further stem-cell-related deaths.

**Hiroshi Ikegaya** *Kyoto Prefectural University of Medicine, Kyoto, Japan.*
*ikegaya@koto.kpu-m.ac.jp*

## Avoid constructing wind farms on peat

Scotland's government is planning to build large-scale wind farms to reduce carbon emissions from electricity production, some of which could be situated on peatlands. We contend that wind farms on peatlands will probably not reduce emissions, unlike those on mineral soils.

Wind farms are often located in upland areas because most of these are windy, distant from residential areas and of low agricultural value. Peatlands are prevalent in UK uplands and are richer in carbon than mineral soils because peats are formed from decomposing wet vegetable matter. Peatlands therefore have a higher net carbon loss when drained for construction.

The UK wind industry uses a method we and our colleagues developed to estimate carbon emissions (D. R. Nayak *et al. Mires Peat* **4,** 9; 2010). On this basis, and assuming current emission factors for electricity generation, our previous work argued that most peatland sites could save on net emissions if peat is not drained and if sites are restored after construction.

However, emissions factors are likely to drop significantly in the future owing to reduced fossil-fuel use in electricity generation (see go.nature.com/lnowou). As a result, peatland sites would be less likely to generate a reduction in carbon emissions, even with careful management. Unless the volume of peat excavated can be significantly reduced relative to energy output, we suggest that construction of wind farms on non-degraded peats should always be avoided.

**Jo Smith, Dali Rani Nayak, Pete Smith** *University of Aberdeen, UK.*
*jo.smith@abdn.ac.uk*

## Improve sanitation on India's railways

A good place to start with India's problems of poor sanitation (see, for example, *Nature* **486,** 185; 2012) would be the country's 150-year-old railway network, which carries 30 million passengers every day. Hygienic sanitation technologies have yet to be installed in all passenger coaches.

The basic lavatory design throws excreta on to the open railway tracks. This system risks spreading pathogens and parasites to distant locations.

One solution would be to install small biogas plants on trains or at stations. These would generate revenue — from excreta — that could be used to employ cleaning and disposal squads.

**Abhishek Sharma,**
**M. K. Unnikrishnan** *Manipal University, Karnataka, India.*
*abhisheksharma0991@gmail.com*
**Ankush Madaan** *McGill University, Montreal, Quebec, Canada.*

# Martin Fleischmann

## (1927–2012)

Pioneering electrochemist who claimed to have discovered cold fusion.

Although a final reckoning should not let genuine achievements be overshadowed by errors, the blot that cold fusion left on Martin Fleischmann's reputation is hard to expunge.

Fleischmann, who died on 3 August at the age of 85 after illness related to Parkinson's disease, heart disease and diabetes, was the first to observe enhanced Raman emission from molecules at surfaces, now the basis of a spectroscopy technique. And he developed ultramicroelectrodes, used as sensitive electrochemical probes.

Nonetheless, he is best known for his claim in 1989 to have initiated nuclear fusion in bench-top apparatus. The 'cold fusion' debacle provoked bitter disputes that reverberate today. Along with polywater and homeopathy, cold fusion is now regarded as one of the most notorious cases of what chemist Irving Langmuir called pathological science: "the science of things that aren't so".

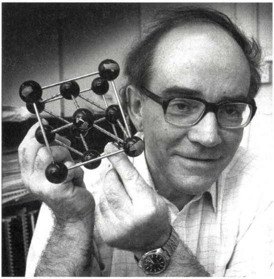
Cold fusion was not really an aberration for Fleischmann, but an extreme example of his willingness to suggest bold and provocative ideas, to take risks and to make imaginative leaps that could sometimes yield a rich harvest.

Fleischmann was born in Carlsbad in Czechoslovakia in 1927. His father was of Jewish heritage, and, just before the German invasion, his family fled to the Netherlands and then to England. Fleischmann studied chemistry at Imperial College London and, after a PhD in electrochemistry, moved to Newcastle University, UK. In 1967 he was appointed as the Faraday Chair of Chemistry at the University of Southampton, UK.

In 1974, Fleischmann and his co-workers observed unusually intense Raman emission (scattered light shifted in energy by interactions with molecular vibrational states) from organic molecules adsorbed on the surface of silver electrodes. Although the enhancement mechanism is still not fully understood, surface-enhanced Raman spectroscopy has become a valuable tool for investigating surface chemistry.

Around 1980, Fleischmann and chemist Mark Wightman independently pioneered the use of ultramicroelectrodes just a few micrometres across, which can be used to study electrode processes that are otherwise inaccessible, for example at low electrolyte concentrations. In 1985, Fleischmann was elected a fellow of Britain's Royal Society.

The cold fusion experiments arose out of Fleischmann's long-standing interest in hydrogen surface chemistry on palladium. Hydrogen molecules adsorbed onto palladium can diffuse into the metal lattice, making palladium a 'sponge' that soaks up large amounts of hydrogen. Very high pressures can build up — perhaps, Fleischmann speculated, high enough to fuse hydrogen nuclei.



THE SOUTHERN DAILY ECHO/SOLENT NEWS

Fleischmann's retirement from Southampton in 1983 freed him to conduct self-funded experiments at the University of Utah in Salt Lake City with his former student Stanley Pons. They electrolysed solutions of lithium deuteroxide, collecting deuterium at the palladium cathode, and claimed to measure more heat output than the energy fed in — a signature, they said, of deuterium fusion within the electrode. One morning, they found that apparatus left running overnight had been vaporized and the fume cupboard destroyed. They believed it was the result of a violent outburst of fusion.

Not until 1989 did Fleischmann, Pons and their student Marvin Hawkins make a move to publish their data. Finding that they were in competition with a team led by physicist Steven Jones at Brigham Young University in Provo, Utah, Fleischmann and Pons initially accused Jones of stealing their ideas. But the groups agreed to coordinate their announcements and to submit papers simultaneously to *Nature* on 24 March 1989. Yet Fleischmann and Pons pre-empted that arrangement, rushing a second paper to the *Journal of Electroanalytical Chemistry*, organizing a press conference on 23 March and faxing their manuscript to *Nature* the same day without telling Jones.

The rest, as they say, is history, told for example in Frank Close's *Too Hot To Handle* (W. H. Allen, 1990). Fleischmann and Pons's announcement shocked the world. Chemists had apparently, at minuscule expense, solved the fusion problem that physicists had been working on for decades. In the attendant flurry, Fleischmann and Pons professed to be too busy to address reviewers' comments and withdrew their *Nature* paper; Jones's account was eventually published (S. E. Jones *et al. Nature* **338,** 737–740; 1989). Despite sporadic claims to the contrary, no comprehensive attempt at replication produced any confirmation of fusion.

Indeed, it was a lack of reproducibility that finally put paid to the cold fusion idea. More bad behaviour followed: Fleischmann refused to describe crucial control experiments; Pons's lawyer threatened to sue a Utah physicist who reported in *Nature* (see M. H. Salamon *et al. Nature* **344,** 401–405; 1990) that he was unable to replicate the work. The University of Utah sought to capitalize on events, throwing US$5 million at a 'National Cold Fusion Institute' that closed two years after it opened.

Fleischmann and Pons moved to France to continue their work with private funding, but later fell out. The biggest casualty of cold fusion was electrochemistry itself, suddenly seeming to be exposed as a morass of charlatanism and poor technique. That was unfair: some of the most authoritative (negative) attempts to replicate the results were conducted by electrochemists.

Fleischmann's tragedy was Shakespearean, not least because he was a sympathetic character: resourceful, energetic, inventive and remembered warmly by collaborators. As Linus Pauling and Fred Hoyle experienced, once you have been proved right against the odds, it becomes harder to accept the possibility of error. To make a mistake or a premature claim, even to fall prey to self-deception, is a risk any scientist runs. The test is how one deals with it. ■

**Philip Ball** *is a writer based in London and was a physical-sciences editor at* Nature *at the time of the cold fusion publications.*
*e-mail: p.ball@btinternet.com*

Principles of Biology
nature EDUCATION
Principles of Biology
Principles of Biology

Principles of **Biology**

by nature **EDUCATION**

Bringing biology to life

To learn more about *Principles of Biology*, the first interactive textbook from Nature Education, visit **www.nature.com/principles**

nature publishing group npg

Affymetrix®

# RNA-SEQ ANALYSIS TAKES WAY TOO LONG

AFFYMETRIX TIME TO INSIGHT
SLOW
FAST

Sequencing is powerful, but it makes genomics harder than it has to be. You need answers now. With Affymetrix® microarrays, get efficient analysis and fast results.

**Down to a Science.**

www.affymetrix.com/downtoascience

# RESEARCH

## NEW ONLINE

*Papers published this week at Nature.com/nature*

**NEWS & VIEWS** Smart 'hygro-responsive' membranes separate oil from water **p.41**

**LETTER** Voyager's slow-motion aerobatics catch the fading solar wind **p.124**

**LETTER** At a stretch these tough new hydrogels could prove extremely useful **p.133**

**PUBLISHED 2 SEPTEMBER 2012**

**CELL BIOLOGY**

### Membrane protein function mapped

Knowledge of membrane protein behaviour is central to an understanding of eukaryotic cell biology. This paper demonstrates that extensive information about membrane proteins of unknown function can be obtained by identifying their interactions with proteins of known function. The hydrophobic nature of membrane protein complexes makes them intractable to conventional affinity purification, but Andrew Emili and colleagues show that soluble membrane complexes from yeast can be affinity-purified in the presence of non-denaturing detergents. They identify co-purifying proteins by mass spectrometry and generate an extensive physical interaction map of membrane protein interactions.

**Interaction landscape of membrane–protein complexes in *Saccharomyces cerevisiae***
M Babu *et al.*
doi:10.1038/nature11354

**STEM CELLS**

### Two-track skin-cell renewal

Skin epidermis consists of a basal layer of proliferative cells over layers of terminally differentiated cells that are progressively enucleated and shed from the skin surface. The cells that maintain this important barrier to infection and injury are generated in the interfollicular epidermis, but whether this involves a single group of stem cells or multiple populations is controversial. Cédric Blanpain and colleagues use lineage tracing and mathematical modelling to show that there are two classes of stem cell in mouse tail skin: a previously reported population of committed progenitors, and slow-cycling stem cells that divide asymmetrically only about four to six times a year.

**Distinct contribution of stem and progenitor cells to epidermal maintenance**
G Mascré *et al.*
doi:10.1038/nature11393

**STRUCTURAL BIOLOGY**

**Structural basis for RNA-duplex recognition and unwinding by the DEAD-box helicase Mss116p**
A L Mallam *et al.*
doi:10.1038/nature11402

**Alternating-access mechanism in conformationally asymmetric trimers of the betaine transporter BetP**
C Perez *et al.*
doi:10.1038/nature11403

**NEUROSCIENCE**

**Nonlinear dendritic processing determines angular tuning of barrel cortex neurons *in vivo***
M Lavzin *et al.*
doi:10.1038/nature11451

**PUBLISHED 5 SEPTEMBER 2012**

**CLIMATE SCIENCE**

### Tropical rain in the air over forests

This global observational analysis demonstrates that forests exert a strong control on rainfall hundreds of kilometres downwind through water cycle feedback. When precipitation falls, some of it returns to the atmosphere via transpiration and evaporation. In the tropics, this process has long been thought to be an important part of the overall precipitation budget. But most evidence has come from model studies, which remain uncertain. Dominick Spracklen and colleagues use remote sensing and atmospheric back trajectory modelling to show that air passage over dense forests produces about twice as much rain as passage over sparse vegetation. They estimate a 12–21% reduction in seasonal precipitation if Amazon deforestation continues at the present rate, and conclude that efforts to curb deforestation are crucial if dramatic impacts on regional rainfall are to be avoided.

**Observations of increased tropical rainfall preceded by air passage over forests**
D V Spracklen *et al.*
doi:10.1038/nature11390

**The rainforest's water pump**
L E O C Aragão
doi:10.1038/nature11485

**ENVIRONMENTAL SCIENCE**

### Seabed smoothed by trawlers

The direct impact of bottom trawling on local fish populations has received much attention, but trawling also affects other aspects of the ocean environment. This paper shows that bottom trawling — a commercial practice in which heavy nets and gear are dragged along the ocean floor — induces sediment reworking and erosion, causing the gradient of the sea floor to become smoother over time. This reduces the morphological complexity of deep-sea environments. The authors draw parallels between the effects of bottom trawling at sea and intensive agriculture on land, with the important difference that on land, ploughing takes place once or twice a year but at sea, bottom trawling can be a frequent occurrence.

**Ploughing the deep sea floor**
P Puig *et al.*
doi:10.1038/nature11410

**INFORMATION TECHNOLOGY**

### Quantum island hopping

Since the first experimental demonstration of quantum teleportation — a central component of quantum communication and quantum computation — physicists have gradually extended the distance over which quantum information is distributed. This study reports long-distance quantum teleportation involving active feed-forward in real time, a feature that will be essential for future applications such as communication between quantum computers. Using two optical links, quantum and classical, quantum teleportation of an independent state was achieved across a 'record' 143 km of free space between the two Canary Islands of La Palma and Tenerife. Significantly, this route bears comparison with the path between a satellite and a ground station.

**Quantum teleportation over 143 kilometres using active feed-forward**
X-S Ma *et al.*
doi:10.1038/nature11472

**DEVELOPMENTAL BIOLOGY**

**A transcriptomic hourglass in plant embryogenesis**
M Quint *et al.*
doi:10.1038/nature11394

MATERIALS SCIENCE

# A hard concept in soft matter

**Hydrogels have many potential applications, but their mechanical strength is low. By simultaneously crosslinking two kinds of polymers in different ways, a highly fracture-resistant hydrogel has been made.** SEE LETTER P.133

KENNETH R. SHULL

The stress required to break a pristine piece of standard window glass is much larger than that required to break a polymer-based acrylic window, yet the acrylic window has a far better chance of surviving an impact with an errant baseball. In fact, the appropriate measure of fracture resistance is not fracture stress, but fracture energy — impact-resistant glass is designed so that the kinetic energy of a baseball is not sufficient to cause catastrophic breakage of the window. Although this concept has been applied quantitatively to relatively stiff materials such as glass, ceramics and metals, our mechanistic understanding of the fracture of soft, highly extensible materials is much more limited. On page 133 of this issue, Sun *et al.*[1] not only address this issue, but also report a highly extensible material that has remarkable mechanical toughness.

The material described by the authors is a hydrogel. Broadly, hydrogels are solutions of a polymer in water, in which the polymer molecules are crosslinked to one another so that the material can support a mechanical load. Because hydrogels consist primarily of water, the concentration of the load-bearing crosslinks is low, and so the mechanical strength of hydrogels is typically also low. Hydrogels are therefore commonly used in applications in which they are not placed under substantial mechanical stress — such as in drug delivery and tissue engineering, in which the role of the gel is to control the distribution of cells or molecular species. Although a solid material is needed for these applications, a material strength of about 1 kilopascal (corresponding to a 10-gram load distributed over an area of 1 square centimetre) is more than sufficient.

The development of much tougher hydrogels, however, would enable a host of other applications to be considered. For example, if hydrogels could be made to withstand physiologically relevant loads, corresponding to about $10^5$ grams over a load-bearing area of a few square centimetres (ref. 2), then it would be possible to prepare materials that mimic the behaviour of cartilage. Such materials would require a compressive strength in the range of several megapascals, and a corresponding fracture toughness that is much larger than any traditional synthetic hydrogel. For a long time, fully synthetic hydrogels with this kind of mechanical strength simply did not exist.

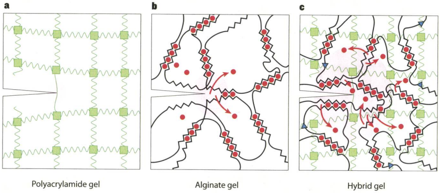


**Figure 1 | Energy dissipation in hydrogels.** When a hydrogel with a notch in it is stretched, crack propagation depends on the polymer chains in the gel. **a**, For covalently crosslinked polymers such as polyacrylamide (green squares indicate crosslinks), the chains ahead of the notch need to break. **b**, In alginate gels — in which calcium ions (red) crosslink binding sites in different chains — the crosslinks ahead of the notch need to break. In both **a** and **b**, the area over which energy is dissipated (pink region) is small, and so cracks propagate easily. **c**, Sun *et al.*[1] report hydrogels that contain crosslinked mixtures of polyacrylamide and alginate (triangles represent crosslinks between different polymer types). They propose that, on stretching, many non-covalent alginate crosslinks break in a wide zone around the head of the notch. Energy is therefore dissipated across a large area, and so the polyacrylamide chains do not break. This makes the gel extremely resistant to crack propagation.

This picture changed significantly in 2003, with the introduction of a set of 'double network' hydrogels[3] that have compressive strengths as high as 20 megapascals, and corresponding fracture energies[4] up to 700 joules per square metre — about 100 times larger than the fracture toughness of a typical hydrogel. These materials are based on two interpenetrating, crosslinked polymer networks. The first network has a fairly high density of covalent crosslinks, giving the gels a modulus (a measure of the material's elastic stiffness) in the megapascal range. This primary network is quite brittle, and it fractures at low applied strains (strain is a measure of the extent to which an object has been deformed by a stress).

The second network has a much lower density of covalent crosslinks than the primary network, and does not contribute substantially to the hydrogel's mechanical properties at low applied strains. But at large strains, such as those encountered in front of a crack that is propagating through the gel, the loosely crosslinked secondary network distributes stress across a relatively narrow 'damage zone'. Energy is dissipated as the primary network is broken into small fragments within this zone, so that the fracture energy of the double network is enormously larger than that of either of the corresponding single networks.

In the previously reported double-network gels[2,3], the fracture energy is dissipated irreversibly as covalent bonds are broken within the damage zone. The materials therefore have excellent fracture toughness, but very poor fatigue life — they behave extremely well during initial compression, but after a subsequent compression the fracture energy is greatly diminished. This limitation is now addressed by Sun and colleagues. The authors replaced the covalently crosslinked primary network with an 'alginate' network that forms non-covalent crosslinks in the presence of calcium ions (Fig. 1). These calcium-based crosslinks form and break reversibly, so that much of the energy that is dissipated when the material is deformed is recoverable. The resulting materials can also be deformed to large strains and yet still retain a high fracture energy. Impressively, some of the gels can be stretched to up to 20 times their original length before fracture occurs, and have corresponding fracture energies of about 9,000 joules per square metre.

Sun and colleagues' materials are noteworthy for three reasons. First, they represent an important extension of the double-network concept, and greatly enhance the maximum extension, fracture energy and retention of material properties of hydrogels during multiple loading cycles. Second, the materials are relatively easy to synthesize compared with previously reported tough hydrogels. And finally, these systems are excellent models for investigating fundamental issues of the fracture behaviour of soft, highly deformable materials.

The Supplementary Information to the paper is full of experimental details that are relevant to this third point. One of the most intriguing results is the implication that, when an existing crack first propagates, energy dissipation is confined to a damage zone that is much smaller than the overall sample size. It is also clear, however, that the new materials can be deformed in such a way that substantial energy is dissipated throughout the entire material before any crack propagation. Additional experiments are needed to sort out the details of energy dissipation in these materials, and the relationship between energy dissipation and crack propagation that forms the core of any investigation of material fracture.

A thorough answer to some of these questions will require a better understanding of the molecular structure of the materials. Sun *et al.* show that the two polymer networks are most probably covalently linked to one another (Fig. 1). The synthesis of materials that do not contain such inter-network links, or in which such links can be introduced at a quantifiable level, is an obvious next step to refine molecular-level models of fracture in soft, highly deformable materials.

Conceptually, the design principles used to produce toughened, 'hard' materials such as window glass are the same as those used to produce tough but 'soft' materials such as polymer gels. Material-specific details matter, however, and different methods are needed to understand the toughening mechanisms in different material classes. Sun and colleagues' gels will certainly motivate continued research by those interested in the mechanical properties of soft materials. The authors have provided some valuable answers about the properties that materials can possess, while at the same time generating a variety of questions for soft-matter scientists to ponder. ■

**Kenneth R. Shull** *is in the Department of Materials Science and Engineering, Northwestern University, Evanston, Illinois 60208-3108, USA.*
*e-mail: k-shull@northwestern.edu*

1. Sun, J.-S. *et al. Nature* **489,** 133–136 (2012).
2. Simon, S. R. *et al. J. Biomech.* **14,** 817–822 (1981).
3. Gong, J. P., Katsuyama, Y., Kurokawa, T. & Osada, Y. *Adv. Mater.* **15,** 1155–1158 (2003).
4. Gong, J. P. *Soft Matter* **6,** 2583–2590 (2010).

---

COSMOLOGY

# The lithium problem

**The theory that predicts how the lightest elements formed after the Big Bang has hitherto failed to explain the amount of cosmic lithium. The detection of interstellar lithium beyond the Milky Way gives this theory a boost.** SEE LETTER P.121

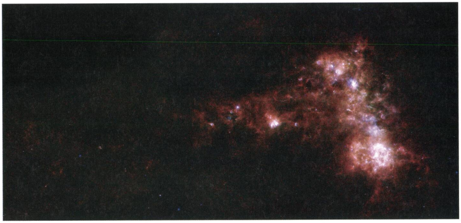
GARIK ISRAELIAN

Our knowledge of the abundances of light elements, such as hydrogen, helium and lithium, in the early Universe has relied on measurements of the chemical content of the atmospheres of old stars in the Milky Way's halo. These observations have long puzzled astronomers because they are in partial disagreement with theoretical predictions, which are based on the Big Bang nucleosynthesis theory and on a precise determination of the cosmic ratio of baryons (particles such as protons and neutrons) to photons. The measured 'primordial' amounts of hydrogen and helium match the predictions, but that of lithium does not. Elsewhere in this issue, Howk and colleagues[1] (page 121) report a measurement of the abundance of the lithium-7 isotope in the interstellar medium of the Small Magellanic Cloud, a dwarf galaxy neighbouring the Milky Way, that is in accord with the Big Bang nucleosynthesis theory.

The nuclei of hydrogen, helium and lithium were created when the Universe was between 2 and 5 minutes old, after the hot primordial plasma had cooled sufficiently for protons and neutrons to form[2]. However, the abundance of lithium is billions of times lower than that of hydrogen and helium. This is because lithium is more prone to being destroyed in stars than hydrogen and helium are, and there are not many processes by which lithium is produced.

Astronomers have long thought that the primordial abundance of lithium is preserved in our Galaxy's stars that are especially old and comparatively cool. Stars have a layered structure. Nuclear-fusion reactions take place in the stars' inner (and hotter) regions but not in their outermost layers. Therefore, the composition of the outermost layers should indicate the chemical content of the matter from which a star has formed. For very old stars, such surface chemical abundances should be close to the primordial values. For younger stars, which formed from material that contained the nuclear-fusion products of previous generations of stars, the surface abundances should be different.

To test the theoretical predictions of the Big Bang nucleosynthesis (BBN) theory, we need to identify astronomical objects in which the primordial abundance values are preserved as much as possible, and we need to account for any remaining influences of chemical evolution. In the early 1980s, astronomers discovered[3] that old, dwarf stars in our Galaxy — Sun-like stars that are poor in metals (elements other than hydrogen and



ESA/NASA/JPL-CALTECH/STSCI

**Figure 1 | The Small Magellanic Cloud.** Howk *et al.*[1] find that the amount of interstellar lithium in the Milky Way's neighbouring Small Magellanic Cloud galaxy is in agreement with the predictions of the Big Bang nucleosynthesis theory. The galaxy is seen here in infrared light collected by the Herschel Space Observatory and the Spitzer Space Telescope.

helium) — share the same lithium abundance irrespective of their temperature and metal content. This 'plateau' was readily interpreted as evidence that the constant lithium abundance was primordial.

Using observations of the cosmic microwave background[4] (relic radiation from the Big Bang) obtained by the Wilkinson Microwave Anisotropy Probe satellite, researchers have been able to make an accurate measurement of the cosmic ratio of baryons to photons. Combined with this measurement, the BBN theory predicts an abundance of the lithium-7 isotope that is about four times that inferred from measurements of old, metal-poor stars in the Milky Way's halo[5]. This mismatch constitutes the 'lithium problem'. The solution to this problem can be sought either by considering modifications to the BBN theory, or by identifying processes by which lithium is destroyed in old, metal-poor halo stars so as to cause the primordial lithium abundance to have evolved, over the stars' lifetimes, to the observed plateau.

An alternative route for tackling the lithium problem — and the one adopted by Howk and colleagues in their study — is to determine the lithium abundance of metal-poor interstellar gas. This approach is unaffected by those processes that can alter the chemical content of stellar atmospheres over time. Howk *et al.* obtained high-quality spectroscopic observations of the lithium spectral line in the metal-poor gas of the Small Magellanic Cloud (Fig. 1). They then derived the total lithium abundance in the galaxy's interstellar medium. This derivation is a difficult task. It requires knowledge of the ionization fraction of lithium and an accurate determination of the amount of lithium locked in interstellar dust grains. The authors used several approaches to account for and measure these quantities. They found that the present-day abundance of interstellar lithium in the Small Magellanic Cloud is almost equal to the BBN predictions.

Howk and colleagues' results are therefore good news for BBN theory. But how can the stellar observations be brought into agreement with the theory? There are many mechanisms that can destroy lithium in stars, all of which imply that the material is processed at temperatures exceeding 2.5 million kelvin. It is, however, difficult to argue that the same mechanism can account for all of the stars that are depleted of lithium. The latest observations[5] of lithium in metal-poor stars in the Galactic halo show a 'meltdown' of the lithium plateau for low metal abundances, such that lithium depletion increases with reduced metal abundance. However, some stars do not follow this trend, and remain on the plateau. This implies that the physics of lithium depletion in metal-poor Sun-like stars is not properly understood.

Magnetic activity, or the presence of a companion star or a giant exoplanet[6], can modify the surface abundance of lithium in Sun-like stars. However, it remains to be investigated whether these factors can explain the lithium content of metal-poor Galactic-halo stars. There are several unanswered questions, but Howk *et al.* provide the first convincing evidence that the lithium abundance in Galactic metal-poor stars is not primordial. ■

**Garik Israelian** *is at the Instituto de Astrofisica de Canarias, La Laguna, 38205 Tenerife, Canary Islands, Spain.*
*e-mail: gil@iac.es*

1. Howk, J. C., Lehner, N., Fields, B. D. & Mathews, G. J. *Nature* **489**, 121–123 (2012).
2. Steigman, G. *Annu. Rev. Nucl. Part. Sci.* **57**, 463–491 (2007).
3. Spite, M. & Spite, F. *Nature* **297**, 483–485 (1982).
4. Dunkley, J. *et al. Astrophys. J. Suppl.* **180**, 306–329 (2009).
5. Bonifacio, P. *et al. Astron. Astrophys.* (in the press); preprint at http://arxiv.org/abs/1204.1641 (2012).
6. Israelian, G. *et al. Nature* **462**, 189–191 (2009).

---

NEUROSCIENCE

# Lessons from heartbreak

**Male fruitflies quickly learn that courting already-mated females is useless. It turns out that a small subset of neurons in the male brain signals this negative experience and controls pheromone sensitivity. SEE LETTER P.145**

AKI EJIMA

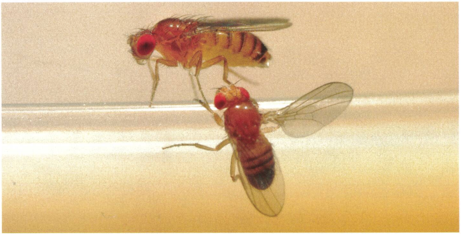
Animals make behavioural decisions on the basis of their prediction of the consequences, and they learn from experience so that they are better prepared for future events. On page 145 of this issue, Keleman *et al.*[1] describe how female rejection enhances the ability of males of the fruitfly *Drosophila melanogaster* to identify a promising mating partner later on.

Courtship by male fruitflies is largely an innate process: the decision to court or not to court depends on the potential mate's scent. A mature virgin female releases aphrodisiac pheromones that trigger the male's courtship, whereas males produce other pheromones that inhibit such behaviour. Moreover, a previously mated female carries some male scent from her previous mating, in addition to her own aphrodisiac pheromones. As a result, naive males court mated females with less enthusiasm than they court virgin females. Nevertheless, once a decision to court is made, the male performs an elaborate courtship ritual (Fig. 1) with no previous instructive experience[2].

There is, however, one aspect of courtship behaviour that can be influenced by experience. In 1979, Siegel and Hall[3] reported that exposing a male fruitfly to a mated female (training) led to suppression of the male's subsequent courtship towards a virgin female (test). Mated females reject courting males because of the influence of sex peptide (SP), a component of the seminal fluid transferred by the previous male during copulation. Rejected males associate the unsuccessful courtship experience with the female's aphrodisiac pheromones, and thus suppress their response to a virgin. This experience-dependent behavioural modification is called courtship conditioning and has served as one of the major experimental paradigms for the study of associative learning in *Drosophila*.

It has become evident, however, that courtship conditioning can follow associative or non-associative mechanisms depending on the nature of the trainer and tester females[4]. For example, it was unclear whether the enhanced courtship suppression produced by repeatedly exposing a male to mated females (at both training and test) was based on associative learning. Because the male is exposed to the same stimuli throughout training and test, the enhanced courtship suppression could be the result of sensory sensitization to negative signals from the mated female.

In fact, a mated female provides two kinds of negative signals to courting males: SP-induced rejection behaviour together with male-derived pheromones such as *cis*-vaccenyl acetate (cVA). To dissociate the effects of these two signal types, Keleman and colleagues used 'pseudomated' females, that is, virgin females that express the SP-encoding gene and therefore reject courting males, but lack cVA pheromone. The authors also allowed SP-deficient males to mate with normal females, which thus became 'pseudovirgins' — mated females that remain receptive (because of the lack of SP) but possess cVA. The authors found that using pseudomated females for training and pseudovirgins for test resulted in the same levels of courtship suppression as using genuine mated females for both training and test. This result indicates that enhanced courtship suppression is a non-associative behavioural modification: a failed copulation attempt (caused by the female's rejection behaviour) enhances the male's sensitivity to cVA, which, in turn, leads



S. ZANKL/NATUREPL.COM

**Figure 1 | Fruitfly courtship.** A male fruitfly (lower) uses his wings in a ritual courtship display to a female.

to exaggerated courtship suppression.

The authors also demonstrated that artificial activation of a specific small subset of dopaminergic neurons — which use dopamine as a neurotransmitter molecule — in the male's brain mimicked courtship training and modified the male's sensitivity to the pheromone. In the fruitfly, dopaminergic neurons are known to have roles in associative learning of tasks linked to odours[5,6], and in male–male courtship conditioning[7] (mature male fruitflies court immature males when first exposed to them, but this behaviour decreases over time as a result of experience). Keleman and colleagues go one step further by uncovering the molecular and cellular mechanisms by which information about a failed courtship experience is signalled through dopaminergic neurons in the male's brain, and how this information affects the male's behavioural sensitivity to cVA.

Male fruitflies have a strong instinct to ingratiate themselves with a potential mate, but the odds are not always on their side. Because a previously fertilized female will not accept a second mating for about a week, males need to know when to pull back. A male-derived pheromone, cVA, helps them to identify such unreceptive females — but why is cVA sensitivity so low in naive males? The authors' finding that courtship training enhances pheromone sensitivity suggests that the male uses his own experience as an indicator of future mating probability. This could help the male to optimize his mating strategy in time and space. It would be interesting to see whether the opposite is also true: does a successful mating experience decrease the male's sensitivity to cVA and, therefore, increase the fly's 'sexual confidence'? ■

**Aki Ejima** *is at the Career-Path Promotion Unit for Young Life Scientists, Kyoto University, Kyoto 606-8501, Japan.*
*e-mail: aki@cp.kyoto-u.ac.jp*

1. Keleman, K. *et al. Nature* **489,** 145–149 (2012).
2. Hall, J. C. *Science* **264,** 1702–1714 (1994).
3. Siegel, R. W. & Hall, J. C. *Proc. Natl Acad. Sci. USA* **76,** 3430–3434 (1979).
4. Griffith, L. C. & Ejima, A. *Learn. Mem.* **16,** 743–750 (2009).
5. Busto, G. U., Cervantes-Sandoval, I. & Davis, R. L. *Physiology* **25,** 338–346 (2010).
6. Liu, C. *et al. Nature* **488,** 512–516 (2012).
7. Neckameyer, W. S. *Learn. Mem.* **5,** 157–165 (1998).

---

CLIMATE CHANGE

# Brief but warm Antarctic summer

**A temperature record derived from measurements of an ice core drilled on James Ross Island, Antarctica, prompts a rethink of what has triggered the recent warming trends on the Antarctic Peninsula.** SEE LETTER P.141

ERIC J. STEIG

In 1842, James Clark Ross sailed past James Ross Island, on the eastern side of the Antarctic Peninsula, and named its high, glaciated volcanic peak, Mount Haddington[1]. In 2008, a team of scientists led by Robert Mulvaney of the British Antarctic Survey successfully drilled an ice core near the summit of Mount Haddington, reaching bedrock at 364 metres below the surface. Now, on page 141 of this issue, Mulvaney and colleagues[2] describe a detailed analysis* of the ice core that allowed them to produce a long record of climate change on the Antarctic Peninsula, one of the fastest-warming regions on Earth. The record stretches back to at least 20,000 years BP (0 yr BP means AD 1950), and may extend to about 50,000 BP.

Much has changed on the Antarctic Peninsula in the 170 years since Ross's voyage there. The most dramatic changes have occurred in just the past two decades, during which a number of large ice shelves have collapsed, altering the geography of the region. Ross's transit along the eastern shore of his namesake island was apparently blocked by ice at the southern entrance to Admiralty Sound (Fig. 1). Only after 1995, with the collapse of the ice shelf in Prince Gustav Channel between James Ross Island and the mainland of the Antarctic Peninsula, did circumnavigation of the island become possible.

Surface melting during unusually warm summers has had a critical role in the recent demise of Antarctic Peninsula ice shelves[3], and it is natural to relate these melting events to anthropogenic global warming[4]. However, most researchers have been reluctant to make this connection, in part because the record of temperature measurements on the peninsula — as elsewhere in Antarctica — is relatively short, and the natural decade-to-decade variability is large[5], making the significance of recent warming trends difficult to assess. Also, it is only on the eastern margin of the Antarctic Peninsula that the summertime temperature trends are large. On the western side, the greatest warming in the past 50 years has occurred in winter and spring[6], as it has in continental West Antarctica[7]. These differing seasonal trends suggest different underlying mechanisms[7], and many studies have attributed the summer warming on the eastern peninsula to atmospheric-circulation change associated with the Antarctic ozone hole in the stratosphere[8] (the atmospheric layer immediately above the troposphere, the lowest portion of the atmosphere). Thus, it has been thought that if human agency has played a part in the demise of Antarctic Peninsula ice shelves, it has primarily been through our destruction of stratospheric ozone rather than through the increased radiative forcing from greenhouse gases in the troposphere.

Mulvaney *et al.* provide a much longer record of temperature than is available from direct instrumental observations. Using the oxygen and hydrogen isotope ratios measured on the

---

*This article and the paper under discussion[2] were published online on 22 August 2012.

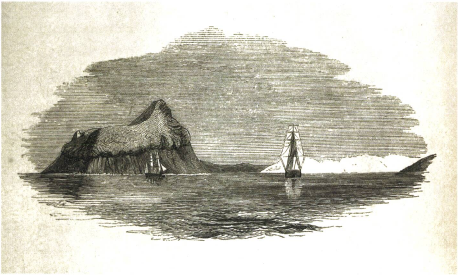


**Figure 1 | Voyage of discovery.** This image from James Clark Ross's *Voyage of Discovery*[1] shows Admiralty Sound blocked by ice in 1842. Cockburn Island is shown on the left, with vessels *HMS Erebus* and *HMS Terror* in the foreground. The edge of James Ross Island is visible on the right. An ice-core temperature record[2] from the summit of James Ross Island shows that recent warming in this area has been unusually rapid.

ice core as palaeothermometers, the authors show that warming began at James Ross Island in the 1920s, well before the advent of chlorofluorocarbon production and the development of the stratospheric ozone hole. This timing is in good agreement with the only long instrumental temperature record available anywhere near the Antarctic Peninsula — on the sub-Antarctic island of Orcadas, some 1,000 kilometres to the northeast[9]. It is also in agreement with instrumental records for the Southern Hemisphere as a whole, and with the ice-core record from the West Antarctic Ice Sheet[10].

Although temperatures on the Antarctic Peninsula comparable to those of the present have certainly occurred in the past, the last time that century-average temperatures were as warm as those of the twentieth to early twenty-first centuries was about 2,000 years ago — corresponding with evidence from marine sediment cores indicating that this was the last time Prince Gustav Channel was open[11]. Thus, the growth and decay of Antarctic Peninsula ice shelves have followed temperature variations over thousands of years.

Mulvaney and colleagues' results provide evidence that the modern occurrence of exceptionally warm temperatures on the Antarctic Peninsula may not be attributable solely either to the decline of stratospheric ozone — the warming trend begins too early — or to natural decadal climate variability. Indeed, one could postulate, as a null hypothesis, that warming on the Antarctic Peninsula is independent of the global-warming trend of the past century. However, the rate of recent warming at James Ross Island is highly unusual, falling within the uppermost 0.3% of all century-scale temperature trends of the past two millennia, which would compel us to reject the null hypothesis with confidence. A caveat is that this conclusion applies only to mean annual temperatures; obtaining seasonal information from ice cores is difficult. These results cannot, therefore, be considered definitive evidence for exceptional long-term trends in summer temperature.

It does not necessarily follow that current warming trends and associated ice-shelf losses will continue. A pivotal influence on Antarctic Peninsula climate, in addition to the effects of greenhouse-gas forcing and ozone changes, are the atmospheric-circulation anomalies that result from climate changes elsewhere, particularly in the tropical Pacific[5,12]. How such anomalies will evolve in the future is highly uncertain[13]. Nevertheless, the unusual temperature increase over the past century suggests that relatively modest radiative forcing from the global increase in greenhouse gases has had a significant effect on the Antarctic Peninsula. Continued increases in both mean annual and summer temperature on the Antarctic Peninsula are a common feature of projections from climate models, given continued increases in greenhouse gases[14]. Mulvaney and colleagues' observations make such projections difficult to dismiss. ■

**Eric J. Steig** *is in the Quaternary Research Center and Department of Earth and Space Sciences, University of Washington, Seattle, Washington 98195, USA.*
*e-mail: steig@uw.edu*

1. Ross, J. C. *A Voyage of Discovery and Research in the Southern and Antarctic Regions, During the Years 1839–43* Vol. 2 (Murray, 1847).
2. Mulvaney, R. *et al. Nature* **489,** 141–144 (2012).
3. Scambos, T. A., Hulbe, C., Fahnestock, M. & Bohlander, J. *J. Glaciol.* **46,** 516–530 (2000).
4. Hodgson, D. A. *Proc. Natl Acad. Sci. USA* **108,** 18859–18860 (2011).

NATURE

## 50 Years Ago

'Antibiotic activity of various types of cannabis resin' — The differences in chemical composition shown by various types of cannabis resin may be explained by the stage of development of a phytochemical process by which cannabidiolic acid is gradually converted to cannabidiol, tetrahydrocannabinols and finally to cannabinol … referred to as 'ripening' of the resin … According to the results obtained, antibiotic activity decreases together with the progress of phytochemical conversion of cannabinols, that is, together with the increase of hashish activity. Antibacterial agent (cannabidiolic acid) is by the ripening process obviously converted into hashish-active constituents (tetrahydrocannabinols). The antibiotically active unripe cannabis seems to be more common in regions having unfavourable climate, whereas tropical samples more often correspond to the ripe, hashish-active drug.
*A. Radošević, M. Kupinić & Lj. Grlić*
**From *Nature* 8 September 1962**

## 100 Years Ago

We are glad to see that progress is gradually being made with the synchronisation of clocks … Last year a committee of the British Science Guild … recommended that, as a beginning, it would probably be well to have a few large public clocks in London synchronised, and that these should be set apart and considered as "standard time clocks." An electric clock which may be used for the purpose suggested by the committee has just been built by the Silent Electric Clock Co. … We understand that this electric clock … is also to be controlled by a master clock directly synchronised from Greenwich. The clock thus represents an up-to-date form of public timekeeper which is likely to be extensively adopted in the future.
**From *Nature* 5 September 1912**

5. Okumura, Y., Schneider, D., Deser, C. & Wilson, R. *J. Clim.* http://dx.doi.org/10.1175/JCLI-D-12-00050.1 (2012).
6. Turner, J. *et al. Int. J. Climatol.* **25,** 279–294 (2005).
7. Steig, E. J. *et al. Nature* **457,** 459–462 (2009).
8. Thompson, D. W. J. *et al. Nature Geosci.* **4,** 741–749 (2011).
9. Zazulie, N., Rusticucci, M. & Solomon, S. *J. Clim.* **23,** 189–196 (2010).
10. Schneider, D. P. & Steig, E. J. *Proc. Natl Acad. Sci. USA* **105,** 12154–12158 (2008).
11. Pudsey, C. J. & Evans, J. *Geology* **29,** 787–790 (2001).
12. Ding, Q., Steig, E. J., Battisti, D. S. & Wallace, J. M. *J. Clim.* http://dx.doi.org/10.1175/JCLI-D-11-00523.1 (2012).
13. Collins, M. *et al. Nature Geosci.* **3,** 391–397 (2010).
14. Bracegirdle, T. J., Connolley, W. M. & Turner, J. *J. Geophys. Res.* **113,** D03103 (2008).

SURFACE SCIENCE

# Separation by reconfiguration

**Membranes have been made that are hygro-responsive — their wetting properties change when immersed in water. This striking property allows the membrane to separate emulsions into their oil and water constituents.**

**ROBERT W. FIELD**

Oil and water don't mix, so the saying goes — unless they form an emulsion, in which case it is difficult to get them apart. Reporting in *Nature Communications*, Tuteja and colleagues[1] describe a simple, scalable method of great potential for separating such 'oily water' mixtures. They have developed membranes whose surfaces are extremely repellent to oil, but which allow water to permeate freely when oily water is filtered through them, so that the retained liquid is principally oil. Unlike mechanical systems such as centrifuges or settling tanks, which separate oil from water only if the oil phase is a distinct dispersion of droplets, the authors' 'smart' membranes separate emulsions highly efficiently. Such hygro-responsive membranes could be developed to clean up oil-contaminated sea water.

Tuteja and colleagues previously reported[2,3] superoleophobic surfaces — ones that resist wetting by liquids that have extremely low surface tension, such as oils and alcohols. The key to making them was the recognition that the surfaces' texture is crucial for superoleophobicity. In particular, re-entrant surface curvature (surfaces that have concave topographic features) is required[2]. So, by making surfaces that have an appropriate chemical composition, roughened texture and re-entrant surface curvature, the authors prepared materials that were extremely resistant to wetting by several liquids. These surfaces can be thought of as omniphobic, because they are highly repellent to water as well as to oils.

More recently, Tuteja's group went further by developing oleophobic membranes[4] that separate oily water emulsions when an electric field is applied across the membrane. This enabled 'on-demand' separation of millilitres of emulsion, but it is questionable whether the system could be used at an industrial scale. Although electrically enhanced processes[5] were an active research area in the 1980s and 1990s, commercial developments have not followed because scaling up is a problem.

The membranes now reported by Tuteja and colleagues[1] are different. The authors describe them as hygro-responsive, a word that derives from the Greek *hygros*, which means wet. This description is certainly pertinent, because wetting of the membranes by water — along with wicking and capillary flow — is vital for their separation properties. The authors prepared their membranes by coating either a stainless-steel mesh or a polyester fabric with a blend of a polymer and an oligomeric material. The resulting non-wetted membranes are both superoleophobic and hydrophobic, but when they are wetted, molecules at the surface of the coating reconfigure in such a way as to enable excellent water permeability while retaining superoleophobicity. This reconfiguration could be attained within a few minutes, which means that the time taken to 'activate' a membrane with water will not be a problem in industrial applications.

A similar reconfiguration has been observed at the surfaces of other polymer films, such as poly(methyl methacrylate), for which the relationship between molecular surface rearrangement and wettability has been well characterized[6]. It has also been noted[7] that surfaces that have been chemically modified by the attachment of amphiphilic macromolecules (polymers that have both hydrophilic and hydrophobic properties) can lead to 'switchable wetting', in which the surface's wetting properties change depending on the properties of fluids to which they are exposed. By taking these materials through several wetting and drying cycles with water, it was shown that surface reconfiguration in these systems is reversible.

Two aspects of the hygro-responsive membranes[1] are particularly striking. First, the water flux through the steel-mesh membrane is exceptionally high at around 43,000 litres per square metre per hour (more than 10 litres per square metre per second). This is more than 1,000 times that of a typical industrial

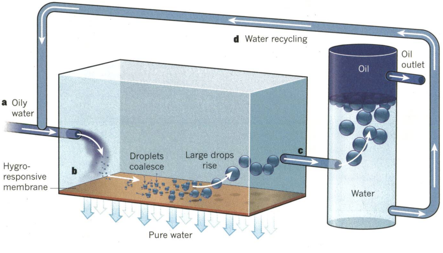


**Figure 1 | Flow scheme for separating oil–water emulsions.** The scheme depicts how Tuteja and colleagues' hygro-responsive membranes[1] might be used in a flow system for separating the constituents of oil–water emulsions. **a,** The oily water is fed into a vessel where it filters through a membrane. Essentially pure water passes through. **b,** The membrane also causes tiny droplets of oil in the emulsion to coalesce at its surface. Once large enough, these rise within the oily water. **c,** A suspension of the large droplets in water is passed into a separate chamber, where the droplets float to the surface and form a separate layer of oil. **d,** The underlying water layer, which still contains a little oil, is recycled back into the flow of oily water for further processing.

ultrafiltration membrane unit. Second, the fact that the authors' technique for making hygro-responsive membranes can be applied to textiles and other surfaces is exciting, because this will enable a range of options to be explored, parallelling the wide range of module types in the membrane industry. Filtration modules based on hygro-responsive membranes, and capable of treating many tonnes of oily water each day, may well emerge soon.

The authors describe their separation technique as a capillary-force-based separation method — that is, one that exploits the difference in capillary forces acting on the individual phases of oily water as it interacts with the membrane. This is a fair description of the process. More questionable is their statement[1] that their process is "solely gravity driven". Although gravity can certainly be used to bring oily water emulsions into contact with the membrane, if an emulsion was pumped between two hygro-responsive membranes, I am confident that water would penetrate through both membranes irrespective of their orientation (and therefore of the influence of gravity). A simple experiment could be performed to test this.

Tuteja and colleagues also provide an equation for the breakthrough pressure of their membranes — the maximum pressure difference across the membrane at which the material prevents the permeation of oil. This enables pumped systems to be designed that use the membranes to separate oil–water emulsions. Such systems would be low-pressure systems in the eyes of process engineers, and would therefore have low operating costs. A design for one possible system is shown in Figure 1.

The authors separated emulsions of water and rapeseed oil as proof of concept of their work. In a related study[8], others have separated mixtures of water and hexadecane (a diesel-like hydrocarbon). However, in the real world, filtration processes suffer from fouling and biofouling of the membranes. Further work using sea water and oil, and a systematic study of possible foulants, should therefore be undertaken to assess the commercial potential of these exciting new membranes. ■

**Robert W. Field** *is in the Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK.*
*e-mail: robert.field@eng.ox.ac.uk*

1. Kota, A. K., Kwon, G., Choi, W., Mabry, J. M. & Tuteja, A. *Nature Commun.* **3,** 1025 (2012).
2. Tuteja, A. *et al. Science* **318,** 1618–1622 (2007).
3. Tuteja, A., Choi, W., Mabry, J. M., McKinley, G. H. & Cohen, R. E. *Proc. Natl Acad. Sci. USA* **105,** 18200–18205 (2008).
4. Kwon, G. *et al. Adv. Mater.* **24,** 3666–3671 (2012).
5. Bowen, W. R. in *Membranes in Bioprocessing* (eds Howell, J. A., Sanchez, V. & Field, R. W.) Ch. 8 (Blackie, Chapman & Hall, 1993).
6. Horinouchi, A., Atarashi, H., Fujii, Y. & Tanaka, K. *Macromolecules* **45,** 4638–4642 (2012).
7. Howarter, J. A., Genson, K. L. & Youngblood, J. P. *ACS Appl. Mater. Interfaces* **3,** 2022–2030 (2011).
8. Howarter, J. A. & Youngblood, J. P. *J. Colloid Interface Sci.* **329,** 127–132 (2009).

---

ASTRONOMY

# Outflows from the first quasars

**Black holes are best known for pulling matter in. But a distant supermassive black hole, observed as it was when the Universe was less than a billion years old, has been seen pushing gas out of its host galaxy.**

DANIEL MORTLOCK

Astronomers have long known of elliptical galaxies, which contain mostly old stars and are largely devoid of interstellar gas. But the finding[1,2] in 2004 that such objects existed about 11 billion years ago, when the Universe was only 3 billion years old, was surprising — it hadn't generally been thought that such galaxies could have formed so early. The most popular explanation[3] was that these ancient ellipticals once hosted the earliest quasars (accreting supermassive black holes), and that the energy released during this quasar phase was sufficient to blow out the galaxy's gas. Maiolino and colleagues[4] now provide a significant boost for the quasar-outflow model in a paper published in *Monthly Notices of the Royal Astronomical Society*. The authors made the remarkable discovery that one such distant quasar, known as SDSS J1148+5251, which is seen as it was when the Universe was less than 1 billion years old, has just the sort of gas outflow required by these models.

The key to this story is the extreme environment at the centre of a galaxy. Most large galaxies, including the Milky Way, harbour at their centres black holes that have roughly a million times the Sun's mass, but in some cases the central black hole can be more than a billion times heavier than the Sun. These black holes are believed to have grown by accreting surrounding gas, a gradual process in which the infalling material is compressed into a disk and heated to such high temperatures that it comfortably outshines all the stars in the host

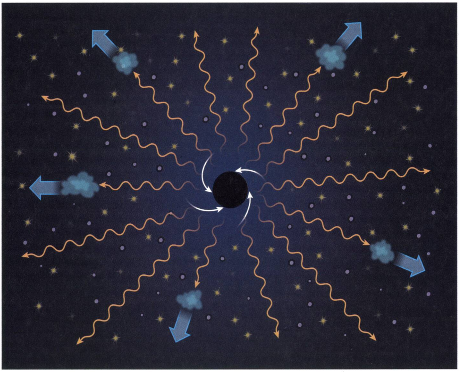


**Figure 1 | Ejection of gas from a galaxy hosting a quasar.** Maiolino *et al.*[4] have found a supermassive black hole (black circle) ejecting gas from its host galaxy. The white arrows show the spiral paths of material being accreted into the black hole, and the orange wavy lines represent photons emitted during this accretion process. Most of the photons escape the galaxy, perhaps to be seen by astronomers, but some impinge on clouds of gas (blue) in the galaxy, and this radiation pressure drives the gas out of the galaxy. The stars (yellow) and dark matter (grey points) are unaffected by the radiation.

galaxy. It is these accreting supermassive black holes that are known as quasars.

Although quasars are generally seen only as unresolved points of light, they have distinctive spectra characterized by broad ultraviolet and optical emission lines, which distinguish them from other astronomical sources. These emission lines are broadened by the Doppler effect that is associated with motion in the environment close to the quasar, revealing the extreme dynamics in the vicinity of the black hole. But the lines reveal little about the motion of the bulk of the interstellar gas farther out in the quasar's host galaxy.

The best way around this problem has been to try to measure emission lines associated with molecules that are not present in the immediate surroundings of the black hole. One possibility is to make observations at submillimetre wavelengths, at which there are several ionized-carbon emission lines. This method has been used to identify outflows from relatively nearby quasars (see, for example, ref. 5). Maiolino *et al.* adopted this approach, using one of the world's most sensitive millimetre arrays, the Institut de Radioastronomie Millimétrique Plateau de Bure Interferometer, to measure the shape — and thus the velocity profile — of an ionized-carbon emission line in the spectrum of SDSS J1148+5251. This light was emitted with a wavelength of 0.158 millimetres but was redshifted by the expansion of the Universe so that it reached Earth with a wavelength of 1.17 millimetres. The data showed not only a core line with a velocity width of a few hundred kilometres per second, as expected of material moving in a large galaxy, but also much broader 'wings' indicative of gas flowing out at speeds of up to 2,000 kilometres per second.

By adopting simple models to describe the geometry of the outflow (which the observations could not reveal), the authors found that the host galaxy of SDSS J1148+5251 was losing 10 solar masses of gas every day. Given that the total molecular-gas content of the galaxy had previously been estimated at 20 billion solar masses[6], the galaxy would have had all of its gas blown out in about 6 million years — a mere instant in cosmological terms. And although the kinetic power of the outflow, some $2 \times 10^{38}$ watts, might seem huge, it is less than 1% of the total power output of the quasar.

Overall, Maiolino and colleagues' data and interpretation paint a coherent picture of gas ejection from quasar host galaxies. However, given that quasars are fuelled by infalling material, it might seem surprising that they can also cause outflows. The explanation is that the light emitted by the quasar exerts a force (termed radiation pressure) on the surrounding gas, and in the extreme situation around a quasar this is strong enough to drive out all of the gas from the galaxy. The stars in the galaxy are so much denser than the gas that they are not noticeably affected, and the non-interacting dark matter between the stars does not experience any radiation pressure at all (Fig. 1).

Maiolino *et al.* also found some evidence that the outflow is visibly extended in their images, which would imply that it spans much of the galaxy. However, the tentative nature of this measurement, and the implication that this would be the largest such outflow ever measured, make this result speculative at best — a point that the authors are careful to make themselves. By contrast, the main finding that quasar SDSS J1148+5251 has been captured in the process of removing gas from its host galaxy seems quite robust, both because of the remarkable data and because of the existence of a compelling theoretical model. ■

**Daniel Mortlock** *is in the Departments of Physics and of Mathematics, Imperial College London, London SW7 2AZ, UK.*
*e-mail: d.mortlock@imperial.ac.uk*

1. Glazebrook, K. *et al. Nature* **430,** 181–184 (2004).
2. Cimatti, A. *et al. Nature* **430,** 184–187 (2004).
3. Silk, J. & Rees, M. J. *Astron. Astrophys.* **331,** L1–L4 (1998).
4. Maiolino, R. *et al. Mon. Not. R. Astron. Soc.* **425,** L66–L70 (2012).
5. Feruglio, C. *et al. Astron. Astrophys.* **518,** L155 (2010).
6. Walter, F. *et al. Nature* **424,** 406–408 (2003).

---

STRUCTURAL BIOLOGY

# A protein engagement RING

**The mechanistic details of the attachment of a small protein, ubiquitin, to other proteins are unclear. Crystal structures of the complexes formed by the E2–ubiquitin and RING E3 enzymes offer new insights.** SEE ARTICLE P.115

**CHRISTOPHER D. LIMA & BRENDA A. SCHULMAN**

Cells use molecular tags to modulate the fates and functions of proteins. One such tag is ubiquitin, a small protein that regulates nearly every facet of cellular function in eukaryotes (organisms such as animals, plants and fungi). Tagging a protein with ubiquitin requires the sequential action of three types of enzyme: E1 activating enzymes attach ubiquitin to a cysteine amino-acid residue on E2 conjugating enzymes, and E3 ligases stimulate ubiquitin transfer from E2–ubiquitin onto a lysine residue of the substrate protein. How E3 enzymes — the most common of which belong to the RING family[1] — carry out the final step has been a long-standing mystery. Now Plechanovová *et al.*[2] (page 115 of this issue) and Dou *et al.*[3] (writing in *Nature Structural & Molecular Biology*) illuminate this mechanism at high resolution, by describing the structures of RING E3 ligases engaged with E2–ubiquitin. Their results suggest a mode of action that could apply to other E3 enzymes.

More than 600 human genes encode RING or RING-like E3 ligases, underscoring their biological importance[1]. Canonical RING proteins contain a zinc-binding region that is rich in cysteine and histidine residues and that, on its own, can bind to E2–ubiquitin and promote ubiquitin transfer[1]. Previous crystal structures revealed some of the interactions between E2 and E3 enzymes, but none of them had captured the elusive association of an E2–ubiquitin intermediate and a RING E3 ligase. This was largely because the link between E2 and ubiquitin is a labile thioester bond.

Plechanovová *et al.* and Dou *et al.* cleverly overcame this challenge by using engineered E2 proteins that were linked to ubiquitin through more-stable bond types (peptide and oxyester bonds, respectively). Both groups of researchers mixed their engineered E2–ubiquitin with an E3 RING ligase, and determined the crystal structures of the resulting RING–E2–ubiquitin protein complexes. For the E3 ligase, Dou *et al.* used a dimeric BIRC7, whereas Plechanovová *et al.* used a tandem protein fusion (RNF4–RNF4) to mimic the RNF4 dimer.

Earlier studies showed that, in the absence of an E3 partner, E2–ubiquitin can adopt many inactive ('open') configurations[4], which presumably prevent the transfer of the molecular ubiquitin tag to a substrate protein (Fig. 1a). The structures determined by Plechanovová *et al.* and Dou *et al.* reveal that RING E3 ligases lock E2–ubiquitin into an activated, closed conformation that is poised for ubiquitin transfer; such a form has also been described in concurrent studies of similar proteins using nuclear magnetic resonance[3,5].

In the RING–E2–ubiquitin crystal structures, certain amino-acid residues of one of

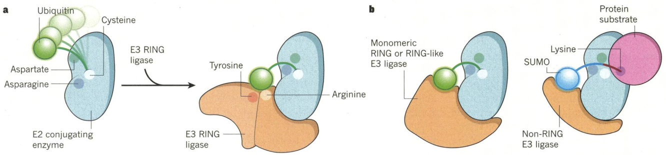


**Figure 1 | A unified model for ubiquitin transfer.** The small protein ubiquitin is attached to a cysteine residue on E2 conjugating enzymes as an intermediate step before being transferred to other proteins through a process that is stimulated by E3 ligase enzymes. **a**, The transfer reaction is presumably hindered by a 'wobbling' movement of ubiquitin when attached to an isolated E2 protein. Plechanovová *et al.*[2] and Dou *et al.*[3] report crystal structures of E2–ubiquitin bound to dimeric E3 ligases of the RING family. They show that RING E3 ligases guide E2–ubiquitin into an active conformation by establishing specific interactions with both ubiquitin and the E2 protein. In particular, an arginine and a tyrosine (or phenylalanine) of the E3 enzyme are crucial for securing ubiquitin into a position that activates transfer. As a result of these interactions, several residues in the E2 protein (such as an asparagine and an aspartate) are reorganized to facilitate the transfer reaction. **b**, Variations on this mechanism are used by monomeric RING, RING-like and some non-RING E3 ligases[9–11] to activate transfer of ubiquitin (or ubiquitin-like proteins such as SUMO) from E2 proteins to a lysine residue on protein substrates.

the two RING monomers interact with both ubiquitin and the E2 protein. Of note, an arginine side chain of one RING monomer bridges the E2 protein and the carboxy-terminal tail of ubiquitin. The opposite RING subunit also contacts ubiquitin through, for example, a highly evolutionarily conserved tyrosine or phenylalanine residue. Moreover, a zinc-bound histidine (which is characteristically found in canonical RING proteins) interacts with ubiquitin through a hydrogen bond.

The crystal structures also show an extensive network of interactions between the E2 protein and its linked ubiquitin. In particular, Plechanovová *et al.* describe a hydrogen bond between a carbonyl oxygen of ubiquitin's C-terminal tail and a highly conserved asparagine side chain of the E2 protein; this asparagine is known[6] to be required for efficient ubiquitin transfer. In addition, an aspartate residue of the E2 protein, which had previously been shown to have a role in activating the substrate protein's lysine[7], is reconfigured in the RING–E2–ubiquitin complexes.

The findings support a model by which RING binding reduces the conformational heterogeneity of E2–ubiquitin and constrains ubiquitin's C-terminal tail in a shallow cleft within the E2 protein (Fig. 1a). As a result, the thioester bond becomes suitably positioned for attack by the substrate protein's lysine, and several residues of the E2 protein are rearranged to promote the transfer reaction. Both groups of authors validated the model through careful biochemical studies. For example, ubiquitin transfer was diminished when the authors made amino-acid changes in the E3 ligase that were predicted to impair its interactions with ubiquitin or with the E2 protein[8]. Moreover, Plechanovová and colleagues describe that their E2–ubiquitin is a competitive inhibitor of E3-mediated ubiquitin transfer to substrate proteins. This result confirms that E2–ubiquitin (in which the two proteins are linked through a peptide bond instead of a thioester) is structurally similar to natural E2–ubiquitin.

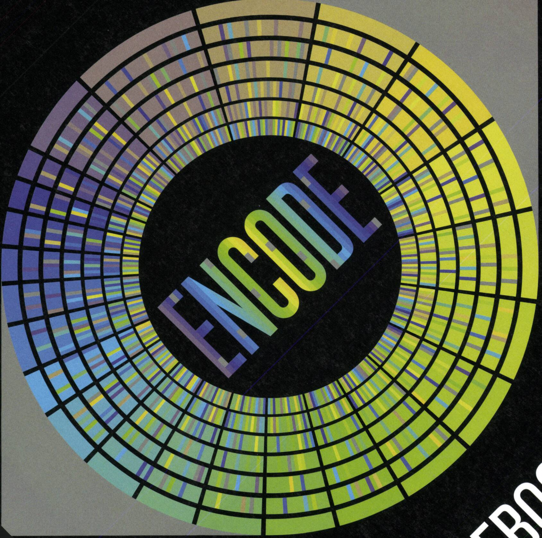
Does the model hold for other E2 and E3 proteins? An earlier study[9] showed that a non-RING E3 ligase (RanBP2) interacts with E2–SUMO in such a way that both E2 and SUMO (a ubiquitin-like protein) are optimally positioned for the transfer reaction to take place. And the RING–E2–ubiquitin structures show striking similarities to that of the protein complex formed by RanBP2, an E2 protein and a SUMO-tagged protein substrate[9] (Fig. 1b). Furthermore, Plechanovová *et al.* show that CHIP, an E3 ligase belonging to the RING-like U-box family, also stimulates ubiquitin transfer by rearranging E2–ubiquitin into a closed configuration. Moreover, computer modelling[10,11] and nuclear magnetic resonance data[5] have indicated that some monomeric RING, or RING-related (SP-RING), E3 ligases contain elements that could lock E2–ubiquitin or E2–SUMO into a closed conformation.

However, there is evidence that, for some E2 proteins, E2–ubiquitin can adopt a closed configuration in the absence of E3 ligases[12–14]. And it is unclear whether some other types of E3 ligase, which transfer ubiquitin through mechanisms different from those used by RING proteins, will follow the model described by the authors. For example, for E3 ligases of the HECT and RBR families, ubiquitin is transferred from an E2 protein onto a cysteine in the E3 enzyme, before being attached to the protein substrate. Although details of the second step await elucidation, it has been reported[15] that HECT binding to E2–ubiquitin promotes tag transfer without stimulating E2–ubiquitin thioester reactivity, in contrast to RING, SP-RING and some other E3 ligases.

In summary, a unified model emerges for those E3 ligases that activate the reactivity of the thioester bond. The binding of an E3 enzyme restricts the conformations available for E2–ubiquitin, which is then forced to adopt a configuration that optimally aligns the thioester for attack by the substrate's lysine. Future studies are required, however, to address how E3–E2–ubiquitin complexes interact with their protein substrates. ■

**Christopher D. Lima** *is in the Structural Biology Program, Sloan-Kettering Institute, New York, New York 10065, USA.*
**Brenda A. Schulman** *is at the Howard Hughes Medical Institute, Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, Tennessee 38105, USA.*
*e-mails: limac@mskcc.org; brenda.schulman@stjude.org*

1. Deshaies, R. J. & Joazeiro, C. A. *Annu. Rev. Biochem.* **78,** 399–434 (2009).
2. Plechanovová, A., Jaffray, E. G., Tatham, M. H., Naismith, J. H. & Hay, R. T. *Nature* **489,** 115–120 (2012).
3. Dou, H., Buetow, L., Sibbet, G. J., Cameron, K. & Huang, D. T. *Nature Struct. Mol. Biol.* http://dx.doi.org/10.1038/nsmb.2379 (2012).
4. Pruneda, J. N., Stoll, K. E., Bolton, L. J., Brzovic, P. S. & Klevit, R. E. *Biochemistry* **50,** 1624–1633 (2011).
5. Pruneda, J. N. *et al. Mol. Cell* http://dx.doi.org/10.1016/j.molcel.2012.07.001 (2012).
6. Wu, P. Y. *et al. EMBO J.* **22,** 5241–5250 (2003).
7. Yunus, A. A. & Lima, C. D. *Nature Struct. Mol. Biol.* **13,** 491–499 (2006).
8. Plechanovová, A. *et al. Nature Struct. Mol. Biol.* **18,** 1052–1059 (2011).
9. Reverter, D. & Lima, C. D. *Nature* **435,** 687–692 (2005).
10. Yunus, A. A. & Lima, C. D. *Mol. Cell* **35,** 669–682 (2009).
11. Dou, H. *et al. Nature Struct. Mol. Biol.* **19,** 184–192 (2012).
12. Hamilton, K. S. *et al. Structure* **9,** 897–904 (2001).
13. Wickliffe, K. E., Lorenz, S., Wemmer, D. E., Kuriyan, J. & Rape, M. *Cell* **144,** 769–781 (2011).
14. Saha, A., Lewis, S., Kleiger, G., Kuhlman, B. & Deshaies, R. J. *Mol. Cell* **42,** 75–83 (2011).
15. Kamadurai, H. B. *et al. Mol. Cell* **36,** 1095–1102 (2009).



# GUIDEBOOK TO
# THE HUMAN GENOME

nature | ENCODE

lumina are very pleased to be part of an im
enome biology, the ENCyclopedia Of DNA Eleme
pitomizes one of the original concepts of the Human G
that a public reference genome sequence would enable sc
draw together all the pieces of biological information associate
fragment of DNA sequence and position them relative to each oth
result, we see the dictionary of genetic words evolving into an encyc
with definitions that provide a foundation for understanding our genome

These studies carry a particular significance for us at Illumina. Using DNA sequence as a single readout for a diversity of biological experiments illustrates the ubiquitous nature of sequence information in biology. The ability to scale these experiments readily across the entire genome is a wonderful demonstration of the power of fast, accurate DNA sequencing technology. These studies re-affirm the importance of continuing to innovate in sequencing technology, to aid researchers worldwide to apply the principles described in these reports to an ever wider range of samples: from whole tissues to single cells.

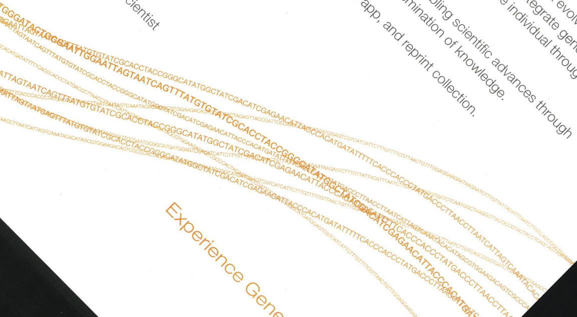
The ability to understand the biological functions encoded by the genome is also of special importance for us as it demonstrates the ability to begin interpreting sequences outside protein coding regions. It opens the way to interpreting individual phenotypic variation that is influenced by sequence variation across the entire genome. As this work evolves it will transform our ability to extract meaning from each person's blueprint, integrate genomic information into medical care and personal lifestyle, and empower the individual through personal genomics in the new era of precision medicine.

Illumina is proud to be a part of the ENCODE project by enabling scientific advances through innovative technology and by sponsoring the public dissemination of knowledge.

Illumina sponsors the micro-site, poster, video, iPad app, and reprint collection.

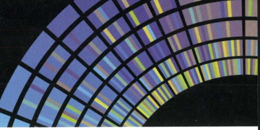
Applauding the project,

David Bentley
Vice President & Chief Scientist
Illumina, Inc.



Experience Gene

# nature | ENCODE



**2001 WILL ALWAYS BE REMEMBERED AS THE YEAR OF THE HUMAN GENOME.** The availability of its sequence transformed biology, and the exemplary way in which hundreds of researchers came together to form a public consortium paved the way for 'big science' in biology. It was an incredible achievement but it was always clear that knowing the 'code' was only the beginning. To understand how cells interpret the information locked within the genome much more needed to be learnt. This became the task of ENCODE, the Encyclopedia Of DNA Elements, the aim of which was to describe all functional elements encoded in the human genome. Nine years after launch, its main efforts culminate in the publication of 30 coordinated papers, 6 of which are in this issue of *Nature*.

Collectively, the papers describe 1,640 data sets generated across 147 different cell types. Among the many important results there is one that stands out above them all: more than 80% of the human genome's components have now been assigned at least one biochemical function.

The implications of the ENCODE findings extend to many fields in biology. In a News & Views Forum on page 52, scientists representing five different areas of research share their views on what the results mean to them and their work. On page 49, Ewan Birney, the leader and coordinator of the ENCODE consortium, discusses the challenges of doing consortium-driven science; related issues are explored in a Careers feature on page 165.

Dizzying amounts of data have been produced by the ENCODE project and are openly accessible; countless more analyses are therefore to be expected, in addition to the multitude now being published. Finding a balance between data collection and analysis is the topic of a News Feature on page 46.

The papers, which are freely available to all, and the articles in this issue are complemented by an extensive range of online features (nature.com/encode). Among them are interactive figures in the overview ENCODE paper, which also features a virtual machine to allow you to interact more closely with the data and their analyses. In line with the community spirit with which the work was undertaken, we also present online the related papers published in *Genome Research* and *Genome Biology*. To help you navigate through the data we have created the Nature ENCODE Explorer and we introduce 'threads', which allow you to explore biological themes between the papers. We hope you enjoy the package.

**Magdalena Skipper** *Senior Editor*
**Ritu Dhand** *Chief Biological Sciences Editor*
**Philip Campbell** *Editor-in-Chief*

## CONTENTS

### FEATURE

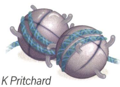
46 **The human encyclopaedia**
*Brendan Maher*

### COMMENT

49 **Lessons for big-data projects**
*Ewan Birney*

### NEWS & VIEWS

52 **ENCODE explained**
*Joseph R Ecker;*
*Wendy A Bickmore;*
*Inês Barroso; Jonathan K Pritchard & Yoav Gilad; Eran Segal*



### ARTICLES

57 **An integrated encyclopedia of DNA elements in the human genome**
*The ENCODE Project Consortium*

75 **The accessible chromatin landscape of the human genome**
*R E Thurman* et al.

83 **An expansive human regulatory lexicon encoded in transcription factor footprints**
*S Neph* et al.

91 **Architecture of the human regulatory network derived from ENCODE data**
*M B Gerstein* et al.

101 **Landscape of transcription in human cells**
*S Djebali* et al.

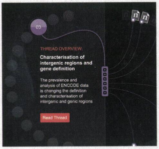
### LETTER

109 **The long-range interaction landscape of gene promoters**
*A Sanyal* et al.

**MORE ONLINE**

**NATURE ENCODE EXPLORER**

Nature ENCODE Explorer offers you a way to explore the wealth of data across all 30 ENCODE papers. By linking relevant paragraphs, figures and tables from the papers, the 'threads' allow you to examine different themes
**nature.com/encode**



THREAD OVERVIEW:

**Characterisation of intergenic regions and gene definition**

The prevalence and analysis of ENCODE data is changing the definition and characterisation of intergenic and genic regions

Available on the App Store

The free Nature ENCODE app for the iPad features all 30 papers plus videos and comment

COVER ILLUSTRATION (PREVIOUS PAGE): CARL DETORRES

# THE HUMAN ENCYCLOPAEDIA

**FIRST THEY SEQUENCED IT. NOW THEY HAVE SURVEYED ITS HINTERLANDS. BUT NO ONE KNOWS HOW MUCH MORE INFORMATION THE HUMAN GENOME HOLDS, OR WHEN TO STOP LOOKING FOR IT.**

BY BRENDAN MAHER

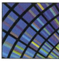
Ewan Birney would like to create a printout of all the genomic data that he and his collaborators have been collecting for the past five years as part of ENCODE, the Encyclopedia of DNA Elements. Finding a place to put it would be a challenge, however. Even if it contained 1,000 base pairs per square centimetre, the printout would stretch 16 metres high and at least 30 kilometres long.

ENCODE was designed to pick up where the Human Genome Project left off. Although that massive effort revealed the blueprint of human biology, it quickly became clear that the instruction manual for reading the blueprint was sketchy at best. Researchers could identify in its 3 billion letters many of the regions that code for proteins, but those make up little more than 1% of the genome, contained in around 20,000 genes — a few familiar objects in an otherwise stark and unrecognizable landscape. Many biologists suspected that the information responsible for the wondrous complexity of humans lay somewhere in the 'deserts' between the genes. ENCODE, which started in 2003, is a massive data-collection effort designed to populate this terrain. The aim is to catalogue the 'functional' DNA sequences that lurk there, learn when and in which cells they are active and trace their effects on how the genome is packaged, regulated and read.

After an initial pilot phase, ENCODE scientists started applying their methods to the entire genome in 2007. Now that phase has come to a close, signalled by the publication of 30 papers, in *Nature*, *Genome Research* and *Genome Biology*. The consortium has assigned some sort of function to roughly 80% of the genome, including more than 70,000 'promoter' regions — the sites, just upstream of genes, where proteins bind to control gene expression — and nearly 400,000 'enhancer' regions that regulate expression of distant genes (see page 57)[1]. But the job is far from done, says Birney, a computational biologist at the European Molecular Biology Laboratory's European Bioinformatics Institute in Hinxton, UK, who coordinated the data analysis for ENCODE. He says that some of the mapping efforts are about halfway to completion, and that deeper characterization of everything the genome is doing is probably only 10% finished. A third phase, now getting under way, will fill out the human instruction manual and provide much more detail.

Many who have dipped a cup into the vast stream of data are excited by the prospect. ENCODE has already illuminated some of the genome's dark corners, creating opportunities to understand how genetic variations affect human traits and diseases. Exploring the myriad regulatory elements revealed by the project and comparing their sequences with those from other mammals promises to reshape scientists' understanding of how humans evolved.

Yet some researchers wonder at what point enough will be enough. "I don't see the runaway train stopping soon," says Chris Ponting, a computational biologist at the University of Oxford, UK. Although Ponting is supportive of the project's goals, he does question whether some aspects of ENCODE will provide a return on the investment, which is estimated to have exceeded US$185 million. But Job Dekker, an ENCODE group leader at the University of Massachusetts Medical School in Worcester, says that realizing ENCODE's potential will require some patience. "It sometimes takes you a long time to know how much can you learn from any given data set," he says.



**ENCODE**
Encyclopedia of DNA Elements
nature.com/encode

Even before the human genome sequence was finished[2], the National Human Genome Research Institute (NHGRI), the main US funder of genomic science, was arguing for a systematic approach to identify functional pieces of DNA. In 2003, it invited biologists to propose pilot projects that would accrue such

## MAKING A GENOME MANUAL

Scientists in the Encyclopedia of DNA Elements Consortium have applied 24 experiment types (across) to more than 150 cell lines (down) to assign functions to as many DNA regions as possible — but the project is still far from complete.

**EXPERIMENTAL TARGETS**

**DNA methylation**: regions layered with chemical methyl groups, which regulate gene expression.

**Open chromatin**: areas in which the DNA and proteins that make up chromatin are accessible to regulatory proteins.

**RNA binding**: positions where regulatory proteins attach to RNA.

**RNA sequences**: regions that are transcribed into RNA.

**ChIP-seq**: technique that reveals where proteins bind to DNA.

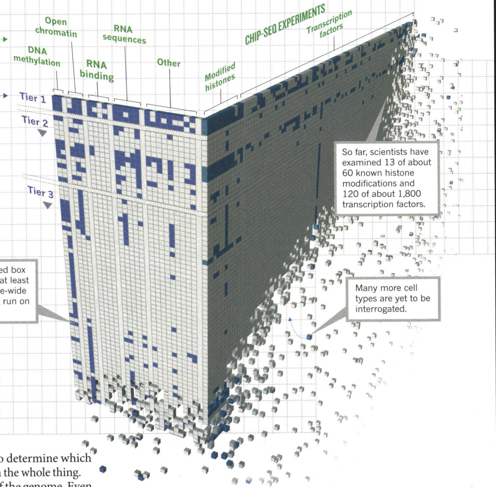
**Modified histones**: histone proteins, which package DNA into chromosomes, modified by chemical marks.

**Transcription factors**: proteins that bind to DNA and regulate transcription.

**CELL LINES**

**Tiers 1 and 2**: widely used cell lines that were given priority.

**Tier 3**: all other cell types.



Every shaded box represents at least one genome-wide experiment run on a cell type.

So far, scientists have examined 13 of about 60 known histone modifications and 120 of about 1,800 transcription factors.

Many more cell types are yet to be interrogated.

information on just 1% of the genome, and help to determine which experimental techniques were likely to work best on the whole thing.

The pilot projects transformed biologists' view of the genome. Even though only a small amount of DNA manufactures protein-coding messenger RNA, for example, the researchers found that much of the genome is 'transcribed' into non-coding RNA molecules, some of which are now known to be important regulators of gene expression. And although many geneticists had thought that the functional elements would be those that are most conserved across species, they actually found that many important regulatory sequences have evolved rapidly. The consortium published its results[3] in 2007, shortly after the NHGRI had issued a second round of requests, this time asking would-be participants to extend their work to the entire genome. This 'scale-up' phase started just as next-generation sequencing machines were taking off, making data acquisition much faster and cheaper. "We produced, I think, five times the data we said we were going to produce without any change in cost," says John Stamatoyannopoulos, an ENCODE group leader at the University of Washington in Seattle.

The 32 groups, including more than 440 scientists, focused on 24 standard types of experiment (see 'Making a genome manual'). They isolated and sequenced the RNA transcribed from the genome, and identified the DNA binding sites for about 120 transcription factors. They mapped the regions of the genome that were carpeted by methyl chemical groups, which generally indicate areas in which genes are silent. They examined patterns of chemical modifications made to histone proteins, which help to package DNA into chromosomes and can signal regions where gene expression is boosted or suppressed. And even though the genome is the same in most human cells, how it is used is not. So the teams did these experiments on multiple cell types — at least 147 — resulting in the 1,648 experiments that ENCODE reports on this week[1,4–8].

Stamatoyannopoulos and his collaborators[4], for example, mapped the regulatory regions in 125 cell types using an enzyme called DNaseI (see page 75). The enzyme has little effect on the DNA that hugs histones, but it chops up DNA that is bound to other regulatory proteins, such as transcription factors. Sequencing the chopped-up DNA suggests where these proteins bind in the different cell types. The team discovered around 2.9 million of these sites altogether. Roughly one-third were found in only one cell type and just 3,700 showed up in all cell types, suggesting major differences in how the genome is regulated from cell to cell.

The real fun starts when the various data sets are layered together. Experiments looking at histone modifications, for example, reveal patterns that correspond with the borders of the DNaseI-sensitive sites. Then researchers can add data showing exactly which transcription factors bind where, and when. The vast desert regions have now been populated with hundreds of thousands of features that contribute to gene regulation. And every cell type uses different combinations and permutations of these features to generate its unique biology. This richness helps to explain how relatively few protein-coding genes can provide the biological complexity necessary to grow and run a human being. ENCODE "is much more than the sum of the parts", says Manolis Kellis, a computational genomicist at the Massachusetts Institute of Technology in Cambridge, who led some of the data-analysis efforts.

The data, which have been released throughout the project, are already helping researchers to make sense of disease genetics. Since 2005, genome-wide association studies (GWAS) have spat out thousands of points on the genome in which a single-letter difference, or variant, seems to be associated with disease risk. But almost 90% of these variants fall outside protein-coding genes, so researchers have little clue as to how they might cause or influence disease.

The map created by ENCODE reveals that many of the disease-linked regions include enhancers or other functional sequences. And cell type is important. Kellis's group looked at some of the variants that are strongly associated with systemic lupus erythematosus, a disease in which the immune system attacks the body's own tissues. The team noticed that the variants identified in GWAS tended to be in regulatory regions of the genome that were active in an immune-cell line, but not necessarily in other types of cell and Kellis's postdoc Lucas Ward has created a web portal called HaploReg, which allows researchers to screen variants identified in GWAS against ENCODE data in a systematic way. "We are now, thanks to ENCODE, able to attack much more complex diseases," Kellis says.

### ARE WE THERE YET?

Researchers could spend years just working with ENCODE's existing data — but there is still much more to come. On its website, the University of California, Santa Cruz, has a telling visual representation of ENCODE's progress: a grid showing which of the 24 experiment types have been done and which of the nearly 180 cell types ENCODE has now examined. It is sparsely populated. A handful of cell lines, including the lab workhorses called HeLa and GM12878, are fairly well filled out. Many, however, have seen just one experiment.

"WE'RE FAR FROM FINISHED. YOU MIGHT ARGUE THAT THIS COULD GO ON FOREVER."

Scientists will fill in many of the blanks as part of the third phase, which Birney refers to as the 'build out'. But they also plan to add more experiments and cell types. One way to do that is to expand the use of a technique known as chromatin immunoprecipitation (ChIP), which looks for all sequences bound to a specific protein, including transcription factors and modified histones. Through a painstaking process, researchers develop antibodies for these DNA binding proteins one by one, use those antibodies to pull the protein and any attached DNA out of cell extracts, and then sequence that DNA.

But at least that is a bounded problem, says Birney, because there are thought to be only about 2,000 such proteins to explore. (ENCODE has already sampled about one-tenth of these.) More difficult is figuring out how many cell lines to interrogate. Most of the experiments so far have been performed on lines that grow readily in culture but have unnatural properties. The cell line GM12878, for example, was created from blood cells using a virus that drives the cells to reproduce, and histones or other factors may bind abnormally to its amped-up genome. HeLa was established from a cervical-cancer biopsy more than 50 years ago and is riddled with genomic rearrangements. Birney recently quipped at a talk that it qualifies as a new species.

ENCODE researchers now want to look at cells taken directly from a person. But because many of these cells do not divide in culture, experiments have to be performed on only a small amount of DNA, and some tissues, such as those in the brain, are difficult to sample. ENCODE collaborators are also starting to talk about delving deeper into how variation between people affects the activity of regulatory elements in the genome. "At some places there's going to be some sequence variation that means a transcription factor is not going to bind here the same way it binds over here," says Mark Gerstein, a computational biologist at Yale University in New Haven, Connecticut, who helped to design the data architecture for ENCODE. Eventually, researchers could end up looking at samples from dozens to hundreds of people.

The range of experiments is expanding, too. One quickly developing area of study involves looking at interactions between parts of the genome in three-dimensional space. If the intervening DNA loops out of the way, enhancer elements can regulate genes hundreds of thousands of base pairs away, so proteins bound to the enhancer can end up interacting with those attached near the gene. Dekker and his collaborators have been developing a technique to map these interactions. First, they use chemicals that fuse DNA-binding proteins together. Then they cut out the intervening loops and sequence the bound DNA, revealing the distant relationships between regulatory elements. They are now scaling up these efforts to explore the interactions across the genome. "This is beyond the simple annotation of the genome. It's the next phase," Dekker says.

The question is, where to stop? Kellis says that some experimental approaches could hit saturation points: if the rate of discoveries falls below a certain threshold, the return on each experiment could become too low to pursue. And, says Kellis, scientists could eventually accumulate enough data to predict the function of unexplored sequences. This process, called imputation, has long been a goal for genome annotation. "I think there's going to be a phase transition where sometimes imputation is going to be more powerful and more accurate than actually doing the experiments," Kellis says.

Yet with thousands of cell types to test and a growing set of tools with which to test them, the project could unfold endlessly. "We're far from finished," says geneticist Rick Myers of the HudsonAlpha Institute for Biotechnology in Huntsville, Alabama. "You might argue that this could go on forever." And that worries some people. The pilot ENCODE project cost an estimated \$55 million; the scale-up was about \$130 million; and the NHGRI could award up to \$123 million in the next phase.

Some researchers argue that they have yet to see a solid return on that investment. For one thing, it has been difficult to collect detailed information on how the ENCODE data are being used. Mike Pazin, a programme director at the NHGRI, has scoured the literature for papers in which ENCODE data played a significant part. He has counted about 300, 110 of which come from labs without ENCODE funding. The exercise was complicated, however, because the word 'encode' shows up in genetics and genomics papers all the time. "Note to self," says Pazin wryly, "make up a unique project name next time around."

A few scientists contacted for this story complain that this isn't much to show from nearly a decade of work, and that the choices of cell lines and transcription factors have been somewhat arbitrary. Some also think that the money eaten up by the project would be better spent on investigator-initiated, hypothesis-driven projects — a complaint that also arose during the Human Genome Project. But unlike the genome project, which had a clear endpoint, critics say that ENCODE could continue to expand and is essentially unfinishable. (None of the scientists would comment on the record, however, for fear that it would affect their funding or that of their postdocs and graduate students.)

Birney sympathizes with the concern that hypothesis-led research needs more funding, but says that "it's the wrong approach to put these things up as direct competition". The NHGRI devotes a lot of its research dollars to big, consortium-led projects such as ENCODE, but it gets just 2% of the total US National Institutes of Health budget, leaving plenty for hypothesis-led work. And Birney argues that the project's systematic approach will pay dividends. "As mundane as these cataloguing efforts are, you've got to put all the parts down on the table before putting it together," he says.

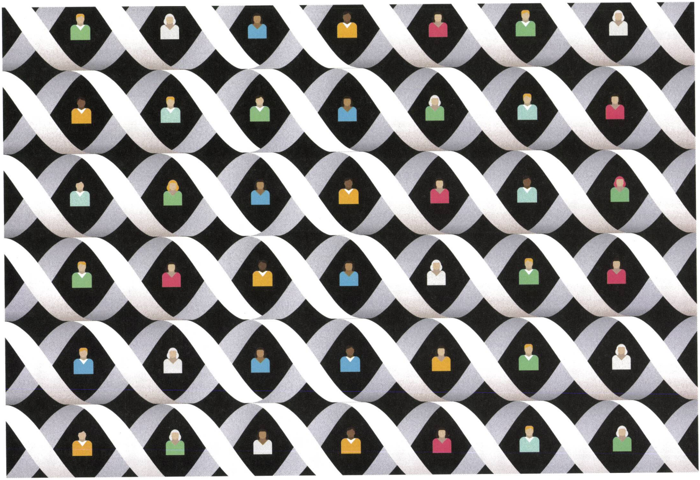
After all, says Gerstein, it took more than half a century to get from the realization that DNA is the hereditary material of life to the sequence of the human genome. "You could almost imagine that the scientific programme for the next century is really understanding that sequence." ■

---

**Brendan Maher** *is a features editor for* Nature.

1. The ENCODE Project Consortium *Nature* **489,** 57–74 (2012).
2. International Human Genome Sequencing Consortium *Nature* **431,** 931–945 (2004).
3. The ENCODE Project Consortium *Nature* **447,** 799–816 (2007).
4. Thurman, R. E. *et al. Nature* **489,** 75–82 (2012).
5. Neph, S. *et al. Nature* **489,** 83–90 (2012).
6. Gerstein, M. B. *et al. Nature* **489,** 91–100 (2012).
7. Djebali, S. *et al. Nature* **489,** 101–108 (2012).
8. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. *Nature* **489,** 109–113 (2012).

THOMAS POROSTOCKY



# Lessons for big-data projects

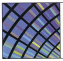
To be successful, consortia need clear management, codes of conduct and participants who are committed to working for the common good, says ENCODE lead analysis coordinator **Ewan Birney**.

The ENCODE consortium has for the past five years been building up an encyclopaedia of functional DNA elements to be used as a reference for the scientific community. Today it publishes 30 publicly accessible papers in three journals — and all are connected to the processed analysis and raw data. This scientific undertaking has inspired new publishing models, such as the interweaving of topic threads between papers in different journals, and will, I hope, have a large impact on biology.

The ENCODE project has delivered an incredible amount of information because of its sheer scale: more than 1,600 experiments on 147 cell types, including 235 antibodies or other assay protocols. The main paper has nearly 450 authors, working from more than 30 institutions.

Because of its complexity (see page 46), the project could not have worked in the same way as one involving just one or two laboratories. Typically, scientists try to do the best science they can, with a limited set of collaborators, to earn grants and publications to do what is best for science, their own careers and their own laboratories.

This mindset doesn't work in consortium science. Instead, researchers must focus on creating the best data set they can. Maybe they will use the data, maybe they won't. What is important is the community resource, not individual success. This requires a shift in perspective to a common goal of data output rather than publications. In turn, the success of consortium participants must be measured at least as much by how their data have enabled science as by the insights they have produced.



**ENCODE**
Encyclopedia of DNA Elements
nature.com/encode

## SUPPORTING THE COMMUNITY

Big-biology consortia such as ENCODE, HapMap and the 1000 Genomes Project approach grand-scale work systematically. For example, they often take a 'catalogue' approach to create foundational resources rather than spotlighting areas of interest, and they use standardized methods, reagents and analysis schemes. The cost of these projects is justified by the breadth of science they support — from genome-wide analysis down to smaller-scale, hypothesis-driven studies. ▶

▶ Has the big project had its day in the current era of 'democratized' data gathering? Certainly the drop in the price of data gathering has changed the game for all biology groups — and nearly always for the better (although there are of course new challenges in how to handle this). But the cheapness of data just extends the reach of large-scale projects; it does not alter the need to create systematic reference data sets. It is hard, if not impossible, to combine smaller data sets into reference data sets — as demonstrated by the initial chromosome maps in the Human Genome Project or the attempt to patch together collections of microarray data into an atlas of gene expression.

Instead, a systematic data 'skeleton' is needed (for genomes, functional elements and variation, for example), around which smaller-scale experiments can add insight, colour and deeper understanding. ENCODE, BLUEPRINT and the 1000 Genomes Project are examples of such skeletons. The main products of ENCODE and similar projects are not just raw data, but also analysed intermediates that allow scientists to choose the level of detail at which they wish to start.

I have been involved in consortia at various levels since 1999. In 2004, I became the coordinator of the ENCODE analysis. I have learned that consortia are difficult to make successful, because they involve people who might be competing with one another in another context. Getting competitors to work openly together towards a shared goal is not trivial. It relies on the good will of all.

ENCODE has made it clear to me that effective consortium science requires all participants to buy into a structure, a code of conduct and the goal of high-quality data that are made accessible and usable to all scientists around the world.

## CLEAR STRUCTURE

In my opinion, for large consortia to succeed, they need to create a structure that is transparent to everyone involved.

This structure cannot follow the classic model of a single institute with a fixed hierarchy, or even a single 'virtual' institute agreed on by multiple partners. Instead, as happened for ENCODE, an open, peer-reviewed process should select and evaluate the partners who are best suited to a self-organized structure. And the structure should be flexible enough to change over time and to encompass multiple sources of funding. Considering each partner as an individual — rather than regarding the consortium as a single group — allows the addition of innovative participants from outside the expected group. ENCODE probably would not have such a great depth of input from statistical groups had the project been funded by a single large grant.

A diverse collection of scientists keeps the ideas fresh and the technology agile. It prevents 'group think'. For example, when there is a shift in technology, labs differ in their uptake. It would be damaging if everyone either committed too early to a poorly performing technology, or delayed uptake of a successful one. Broad participation also connects the output to a much larger audience worldwide.

Large consortia do, however, need to avoid a common pitfall: sharing the responsibility between too many principal investigators and senior postdoctoral fellows. This renders decision-making difficult. Without a core structure, there is a risk that members will shift their focus to their own interest areas at the expense of the overall project. At the same time, these projects are too big and complex to be managed by one person, who is unlikely to have expertise in all the relevant areas. Initiatives that are piloted by one or a few principal investigators are more common in consortia working on diseases, and in my experience they often lack an operational project manager with a well-defined role.

> ***"Consortium science involves interaction between humans, with all the social complications this entails."***

The ENCODE consortium had an internal structure that I believe was instrumental to its success. It had a 'spine' of leadership comprising: scientifically aware project officers in the primary funding agency, the National Human Genome Research Institute at the US National Institutes of Health; a few leading scientists with goals aligned to the consortium; and one or two scientific project managers hired inside the consortium who had a detailed understanding of all the tasks and people involved. ENCODE's two key project coordinators (Ian Dunham and Anshul Kundaje) were funded for the lifetime of the project through a grant for which I was the principal investigator. Successful consortia tend to have similar core structures, suggesting that this is a natural and effective way to organize such projects.

The spine was able to resolve some of the most complex problems — both scientific and social — such as sorting out a quality-control disagreement between a data-production and data-analysis group. As in any endeavour that involves many individuals, communication channels are crucial for success. We should have explicitly broadcast the existence of this spine both to the group and externally, to provide more transparency with respect to how decisions were made.

I also think that funding agencies should become more involved in shaping consortia. They should be flexible enough to shift their support from one group to another as needed, with adequate warning, and to withdraw funding from poorly performing or uncooperative partners — again with warning and with real consequences. Funding agreements often include such terms and conditions, but they are rarely used, perhaps because the threat of action is enough. And perhaps funding agencies feel uncomfortable, understandably, taking on such a scientifically directive role. But the responsibility for the overall success of the project rests firmly with the funding agency, so it must feel empowered to intervene when necessary.

## CODES OF CONDUCT

Consortium science involves interaction between humans, with all the social complications this entails. It happens across multiple sites and time zones, and the partners generally communicate electronically, rather than in person. Misunderstandings and clashes can arise because of cultural differences — at national, organizational and individual levels.

To ensure that things run smoothly, rules are essential. An agreed-upon, written and publicly accessible code of conduct is extremely beneficial to large consortia, particularly when they need to incorporate less-experienced partners. ENCODE had several written rules, on issues such as data release, and these were circulated internally.

Such rules help to ensure that partners work within the goals of the consortium and do not (consciously or unconsciously) form a cartel that controls access to the data and analysis. An advisory board should regularly scrutinize internal and external partners for scientific impact, capacity to deliver and ability to interact effectively. Although I am confident that ENCODE did not restrict access to data or analysis through the rules of the funding agency, outside groups occasionally had that impression, and that is a failing I deeply regret.

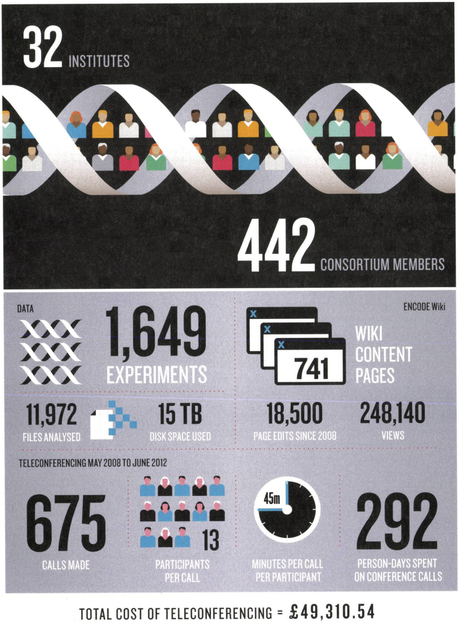
We should also have had written guidelines on how to transfer work between groups, how to assign credit when papers are published and how and when project officers should communicate, especially during times of conflict. Implicit rules of behaviour in consortia are often exploited by more experienced participants.

Large consortia clearly benefit from an open-door policy that allows new, unfunded analysts to participate. And when these individuals join the group or work with released consortium data, their analyses should be considered equally creditable and stigma-free relative to those performed by long-standing group members.

That brings us to error-catching. Big projects generate errors and have a range

## BY THE NUMBERS

The ENCODE project involved hundreds of people from around the world, and a lot of editing, disk space and phone calls.


32 INSTITUTES
442 CONSORTIUM MEMBERS
DATA
1,649 EXPERIMENTS
ENCODE Wiki
741 WIKI CONTENT PAGES
11,972 FILES ANALYSED
15 TB DISK SPACE USED
18,500 PAGE EDITS SINCE 2008
248,140 VIEWS
TELECONFERENCING MAY 2008 TO JUNE 2012
675 CALLS MADE
13 PARTICIPANTS PER CALL
45m MINUTES PER CALL PER PARTICIPANT
292 PERSON-DAYS SPENT ON CONFERENCE CALLS
TOTAL COST OF TELECONFERENCING = £49,310.54

THOMAS POROSTOCKY; SOURCE: MEETINGZONE

of artefacts, so most researchers agree that data should be released to the larger community sooner rather than later. In ENCODE, we came to understand how time-consuming and involved quality control at scale is. It was not until around halfway through the process that we were able to assess the experiments retrospectively with a formalized, centralized, quality-control system. Most experiments were exemplary; some had to be redone. A few had to be left out.

The quality-control metrics and our final 'call' on whether a data set would be in or out is publicly accessible on the project website. Although important and biologically correct, some experiments scored low on quality-control metrics because they had, for example, very few true sites where a protein bound to DNA. Other sources of error, such as that from a cross-reactive antibody, generated excellent scores — the antibody 'worked' because it bound to a particular class of molecule, but it also bound to many others that were not predicted by the analysis. I wish now that we had accelerated the centralized quality-control process earlier, and been more open about this process.

Although most errors are caught within a consortium before they are released, new analysis of public data inevitably uncovers more, particularly early in data production. When analysing such early data, external groups should report such errors promptly and without rancour. Although funders need to measure data quality in a standardized way, during early data production consortia should really be judged not by absolute error rates, but by how quickly they can rectify reported errors.

Funders have considerable influence in how raw and analysed data are released, and should design policies that maximize reuse. Early data-release policies focused on how data should be shared before publication, with clumsy etiquette-based restrictions on the first publications of global analysis, such as waiting for the authors who generated the data to publish their analyses before others can publish on the entire data set. These agreements are starting to show their age and a lack of clarity.

The new era of analysis calls for a rethink, with more focus on the release of intermediate analysis throughout the project, so that the community can use the resource more fully during the project; the 1000 Genomes consortium has done well in this regard.

### DOES IT DELIVER?

The overall importance of consortia science can not be assessed until years after the data are assembled. But reference data sets are repeatedly used by numerous scientists worldwide, often long after the consortium disbands. We already know of more than 100 non-consortium publications that make use of ENCODE data, and I expect many more in the forthcoming years.

Even if massive projects are successful, I feel strongly that the vast majority of funding should still go to smaller, more creative, hypothesis-led science.

For consortium participants, my call for more scrutiny, more clarity and more independent utilization of the data might seem restrictive, but I am confident that it will only benefit science and scientists in the long run. Even if large consortia receive only a small proportion of a discipline's funding, that can be a substantial amount when concentrated on a limited set of groups. If this is to continue, the entire community must be able to understand and use the resultant data.

ENCODE is a foundational data set for understanding the human genome. I am proud of what we have delivered, but there are things we could have done better. I hope that other groups can learn from our experience. ■

**Ewan Birney** *is lead ENCODE analysis coordinator and associate director of the European Molecular Biology Laboratory's European Bioinformatics Institute in Hinxton, UK.*
*e-mail: birney@ebi.ac.uk*

**FORUM:** Genomics

# ENCODE explained

**The Encyclopedia of DNA Elements (ENCODE) project dishes up a hearty banquet of data that illuminate the roles of the functional elements of the human genome. Here, six scientists describe the project and discuss how the data are influencing research directions across many fields.** SEE ARTICLES P.57, P.75, P.83, P.91, P.101 & LETTER P.109

## Serving up a genome feast

**JOSEPH R. ECKER**

Starting with a list of simple ingredients and blending them in the precise amounts needed to prepare a gourmet meal is a challenging task. In many respects, this task is analogous to the goal of the ENCODE project[1], the recent progress of which is described in this issue[2–7]. The project aims to fully describe the list of common ingredients (functional elements) that make up the human genome (Fig. 1). When mixed in the right proportions, these ingredients constitute the information needed to build all the types of cells, body organs and, ultimately, an entire person from a single genome.

The ENCODE pilot project[8] focused on just 1% of the genome — a mere appetizer — and its results hinted that the list of human genes was incomplete. Although there was scepticism about the feasibility of scaling up the project to the entire genome and to many hundreds of cell types, recent advances in low-cost, rapid DNA-sequencing technology radically changed that view[9]. Now the ENCODE consortium presents a menu of 1,640 genome-wide data sets prepared from 147 cell types, providing a six-course serving of papers in *Nature*, along with many companion publications in other journals.

One of the more remarkable findings described in the consortium's 'entrée' paper (page 57)[2] is that 80% of the genome contains elements linked to biochemical functions, dispatching the widely held view that the human genome is mostly 'junk DNA'. The authors report that the space between genes is filled with enhancers (regulatory DNA elements), promoters (the sites at which DNA's transcription into RNA is initiated) and numerous previously overlooked regions that encode RNA transcripts that are not translated into proteins but might have regulatory roles. Of note, these results show that many DNA variants previously correlated with certain diseases lie within or very near non-coding functional DNA elements, providing new leads for linking genetic variation and disease.

The five companion articles[3–7] dish up diverse sets of genome-wide data regarding the mapping of transcribed regions, DNA binding of regulatory proteins (transcription factors) and the structure and modifications of chromatin (the association of DNA and proteins that makes up chromosomes), among other delicacies.

Djebali and colleagues[3] (page 101) describe ultra-deep sequencing of RNAs prepared from many different cell lines and from specific compartments within the cells. They conclude that about 75% of the genome is transcribed at some point in some cells, and that genes are highly interlaced with overlapping transcripts that are synthesized from both DNA strands. These findings force a rethink of the definition of a gene and of the minimum unit of heredity.

Moving on to the second and third courses, Thurman *et al.*[4] and Neph *et al.*[5] (pages 75 and 83) have prepared two tasty chromatin-related treats. Both studies are based on the DNase I hypersensitivity assay, which detects genomic regions at which enzyme access to, and subsequent cleavage of, DNA is unobstructed by chromatin proteins. The authors identified cell-specific patterns of DNase I hypersensitive sites that show remarkable concordance with experimentally determined and computationally predicted binding sites of transcription factors. Moreover, they have doubled the number of known recognition sequences for DNA-binding proteins in the human genome, and have revealed a 50-base-pair 'footprint' that is present in thousands of promoters[5].

The next course, provided by Gerstein and colleagues[6] (page 91) examines the principles behind the wiring of transcription-factor networks. In addition to assigning relatively simple functions to genome elements (such as 'protein X binds to DNA element Y'), this study attempts to clarify the hierarchies of transcription factors and how the intertwined networks arise.

**ENCODE**
Encyclopedia of DNA Elements
nature.com/encode

Beyond the linear organization of genes and transcripts on chromosomes lies a more complex (and still poorly understood) network of chromosome loops and twists through which promoters and more distal elements, such as enhancers, can communicate their regulatory information to each other. In the final course of the ENCODE genome feast, Sanyal and colleagues[7] (page 109) map more than 1,000 of these long-range signals in each cell type. Their findings begin to overturn the long-held (and probably oversimplified) prediction that the regulation of a gene is dominated by its proximity to the closest regulatory elements.

> ***"These findings force a rethink of the definition of a gene and of the minimum unit of heredity."***

One of the major future challenges for ENCODE (and similarly ambitious projects) will be to capture the dynamic aspects of gene regulation. Most assays provide a single snapshot of cellular regulatory events, whereas a time series capturing how such processes change is preferable. Additionally, the examination of large batches of cells — as required for the current assays — may present too simplified a view of the underlying regulatory complexity, because individual cells in a batch (despite being genetically identical) can sometimes behave in different ways. The development of new technologies aimed at the simultaneous capture of multiple data types, along with their regulatory dynamics in single cells, would help to tackle these issues.

A further challenge is identifying how the genomic ingredients are combined to assemble the gene networks and biochemical pathways that carry out complex functions, such as cell-to-cell communication, which enable organs and tissues to develop. An even greater challenge will be to use the rapidly growing body

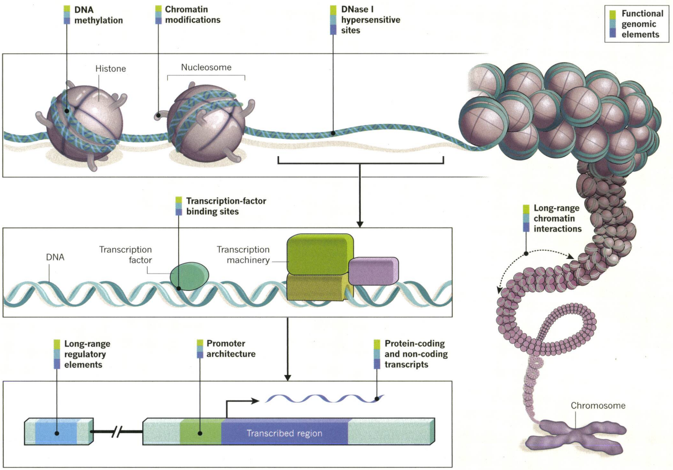

DNA methylation
Chromatin modifications
DNase I hypersensitive sites
Functional genomic elements
Histone
Nucleosome
Transcription-factor binding sites
DNA
Transcription factor
Transcription machinery
Long-range chromatin interactions
Long-range regulatory elements
Promoter architecture
Protein-coding and non-coding transcripts
Transcribed region
Chromosome

**Figure 1 | Beyond the sequence.** The ENCODE project[2–7] provides information on the human genome far beyond that contained within the DNA sequence — it describes the functional genomic elements that orchestrate the development and function of a human. The project contains data about the degree of DNA methylation and chemical modifications to histones that can influence the rate of transcription of DNA into RNA molecules (histones are the proteins around which DNA is wound to form chromatin). ENCODE also examines long-range chromatin interactions, such as looping, that alter the relative proximities of different chromosomal regions in three dimensions and also affect transcription. Furthermore, the project describes the binding activity of transcription-factor proteins and the architecture (location and sequence) of gene-regulatory DNA elements, which include the promoter region upstream of the point at which transcription of an RNA molecule begins, and more distant (long-range) regulatory elements. Another section of the project was devoted to testing the accessibility of the genome to the DNA-cleavage protein DNase I. These accessible regions, called DNase I hypersensitive sites, are thought to indicate specific sequences at which the binding of transcription factors and transcription-machinery proteins has caused nucleosome displacement. In addition, ENCODE catalogues the sequences and quantities of RNA transcripts, from both non-coding and protein-coding regions.

of data from genome-sequencing projects to understand the range of human phenotypes (traits), from normal developmental processes, such as ageing, to disorders such as Alzheimer's disease[10].

Achieving these ambitious goals may require a parallel investment of functional studies using simpler organisms — for example, of the type that might be found scampering around the floor, snatching up crumbs in the chefs' kitchen. All in all, however, the ENCODE project has served up an all-you-can-eat feast of genomic data that we will be digesting for some time. Bon appétit!

---

**Joseph R. Ecker** *is at the Howard Hughes Medical Institute and the Salk Institute for Biological Studies, La Jolla, California 92037, USA.*
*e-mail: ecker@salk.edu*

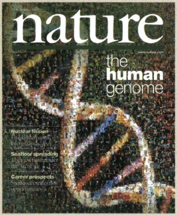
# Expression control

**WENDY A. BICKMORE**

Once the human genome had been sequenced, it became apparent that an encyclopaedic knowledge of chromatin organization would be needed if we were to understand how gene expression is regulated. The ENCODE project goes a long way to achieving this goal and highlights the pivotal role of transcription factors in sculpting the chromatin landscape.

Although some of the analyses largely confirm conclusions from previous smaller-scale studies, this treasure trove of genome-wide data provides fresh insight into regulatory pathways and identifies prodigious numbers of regulatory elements. This is particularly so for Thurman and colleagues' data[4] regarding DNase I hypersensitive sites (DHSs) and for Gerstein and colleagues' results[6] concerning DNA binding of transcription factors. DHSs are genomic regions that are accessible to enzymatic cleavage as a result of the displacement of nucleosomes (the basic units of chromatin) by DNA-binding proteins (Fig. 1). They are the hallmark of cell-type-specific enhancers, which are often located far away from promoters.

The ENCODE papers expose the profusion of DHSs — more than 200,000 per cell type, far outstripping the number of promoters — and their variability between cell types. Through the simultaneous presence in the same cell type of a DHS and a nearby active promoter, the researchers paired half a million enhancers with their probable target genes. But this leaves



## 11 Years Ago

## The draft human genome

**OUR GENOME UNVEILED**

Unless the human genome contains a lot of genes that are opaque to our computers, it is clear that we do not gain our undoubted complexity over worms and plants by using many more genes. Understanding what does give us our complexity — our enormous behavioural repertoire, ability to produce conscious action, remarkable physical coordination (shared with other vertebrates), precisely tuned alterations in response to external variations of the environment, learning, memory … need I go on? — remains a challenge for the future.

*David Baltimore*

**From *Nature* 15 February 2001**

**GENOME SPEAK**

With the draft in hand, researchers have a new tool for studying the regulatory regions and networks of genes. Comparisons with other genomes should reveal common regulatory elements, and the environments of genes shared with other species may offer insight into function and regulation beyond the level of individual genes. The draft is also a starting point for studies of the three-dimensional packing of the genome into a cell's nucleus. Such packing is likely to influence gene regulation … The human genome lies before us, ready for interpretation.

*Peer Bork and Richard Copley*

**From *Nature* 15 February 2001**

more than 2 million putative enhancers without known targets, revealing the enormous expanse of the regulatory genome landscape that is yet to be explored. Chromosome-conformation-capture methods that detect long-range physical associations between distant DNA regions are attempting to bridge this gap. Indeed, Sanyal and colleagues[7] applied these techniques to survey such associations across 1% of the genome.

The ENCODE data start to paint a picture of the logic and architecture of transcriptional networks, in which DNA binding of a few high-affinity transcription factors displaces nucleosomes and creates a DHS, which in turn facilitates the binding of further, lower-affinity factors. The results also support the idea that transcription-factor binding can block DNA methylation (a chemical modification of DNA that affects gene expression), rather than the other way around — which is highly relevant to the interpretation of disease-associated sites of altered DNA methylation[11].

The exquisite cell-type specificity of regulatory elements revealed by the ENCODE studies emphasizes the importance of having appropriate biological material on which to test hypotheses. The researchers have focused their efforts on a set of well-established cell lines, with selected assays extended to some freshly isolated cells. Challenges for the future include following the dynamic changes in the regulatory landscape during specific developmental pathways, and understanding chromatin structure in tissues containing heterogeneous cell populations.

**Wendy A. Bickmore** *is in the Medical Research Council Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK.*
*e-mail: wendy.bickmore@igmm.ed.ac.uk*

# Non-coding but functional

INÊS BARROSO

The vast majority of the human genome does not code for proteins and, until now, did not seem to contain defined gene-regulatory elements. Why evolution would maintain large amounts of 'useless' DNA had remained a mystery, and seemed wasteful. It turns out, however, that there are good reasons to keep this DNA. Results from the ENCODE project[2–8] show that most of these stretches of DNA harbour regions that bind proteins and RNA molecules, bringing these into positions from which they cooperate with each other to regulate the function and level of expression of protein-coding genes. In addition, it seems that widespread transcription from non-coding DNA potentially acts as a reservoir for the creation of new functional molecules, such as regulatory RNAs.

What are the implications of these results for genetic studies of complex human traits and disease? Genome-wide association studies (GWAS), which link variations in DNA sequence with specific traits and diseases, have in recent years become the workhorse of the field, and have identified thousands of DNA variants associated with hundreds of complex traits (such as height) and diseases (such as diabetes). But association is not causality, and identifying those variants that are causally linked to a given disease or trait, and understanding how they exert such influence, has been difficult. Furthermore, most of these associated variants lie in non-coding regions, so their functional effects have remained undefined.

> ***"The results imply that sequencing studies focusing on protein-coding sequences risk missing crucial parts of the genome."***

The ENCODE project provides a detailed map of additional functional non-coding units in the human genome, including some that have cell-type-specific activity. In fact, the catalogue contains many more functional non-coding regions than genes. These data show that results of GWAS are typically enriched for variants that lie within such non-coding functional units, sometimes in a cell-type-specific manner that is consistent with certain traits, suggesting that many of these regions could be causally linked to disease. Thus, the project demonstrates that non-coding regions must be considered when interpreting GWAS results, and it provides a strong motivation for reinterpreting previous GWAS findings. Furthermore, these results imply that sequencing studies focusing on protein-coding sequences (the 'exome') risk missing crucial parts of the genome and the ability to identify true causal variants.

However, although the ENCODE catalogues represent a remarkable tour de force, they contain only an initial exploration of the depths of our genome, because many more cell types must yet be investigated. Some of the remaining challenges for scientists searching for causal disease variants lie in: accessing data derived from cell types and tissues relevant to the disease under study; understanding how these functional units affect genes that may be distantly located[7]; and the ability to generalize such results to the entire organism.

**Inês Barroso** *is at the Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK, and at the University of Cambridge Metabolic Research Laboratories and NIHR Cambridge Biomedical Research Centre, Cambridge, UK.*
*e-mail: ib1@sanger.ac.uk*

## Evolution and the code

**JONATHAN K. PRITCHARD & YOAV GILAD**

One of the great challenges in evolutionary biology is to understand how differences in DNA sequence between species determine differences in their phenotypes. Evolutionary change may occur both through changes in protein-coding sequences and through sequence changes that alter gene regulation.

There is growing recognition of the importance of this regulatory evolution, on the basis of numerous specific examples as well as on theoretical grounds. It has been argued that potentially adaptive changes to protein-coding sequences may often be prevented by natural selection because, even if they are beneficial in one cell type or tissue, they may be detrimental elsewhere in the organism. By contrast, because gene-regulatory sequences are frequently associated with temporally and spatially specific gene-expression patterns, changes in these regions may modify the function of only certain cell types at specific times, making it more likely that they will confer an evolutionary advantage[12].

However, until now there has been little information about which genomic regions have regulatory activity. The ENCODE project has provided a first draft of a 'parts list' of these regulatory elements, in a wide range of cell types, and moves us considerably closer to one of the key goals of genomics: understanding the functional roles (if any) of every position in the human genome.

Nonetheless, it will take a great deal of work to identify the critical sequence changes in the newly identified regulatory elements that drive functional differences between humans and other species. There are some precedents for identifying key regulatory differences (see, for example, ref. 13), but ENCODE's improved identification of regulatory elements should greatly accelerate progress in this area. The data may also allow researchers to begin to identify sequence alterations occurring simultaneously in multiple genomic regions, which, when added together, drive phenotypic change — a process called polygenic adaptation[14].

However, despite the progress brought by the ENCODE consortium and other research groups, it remains difficult to discern with confidence which variants in putative regulatory regions will drive functional changes, and what these changes will be. We also still have an incomplete understanding of how regulatory sequences are linked to target genes. Furthermore, the ENCODE project focused mainly on the control of transcription, but many aspects of post-transcriptional regulation, which may also drive evolutionary changes, are yet to be fully explored.

Nonetheless, these are exciting times for studies of the evolution of gene regulation. With such new resources in hand, we can expect to see many more descriptions of adaptive regulatory evolution, and how this has contributed to human evolution.

---

**Jonathan K. Pritchard** *and* **Yoav Gilad** *are in the Department of Human Genetics, University of Chicago, Chicago 60637 Illinois, USA.* **J.K.P.** *is also at the Howard Hughes Medical Institute, University of Chicago.*
*e-mails: pritch@uchicago.edu; gilad@uchicago.edu*

## From catalogue to function

**ERAN SEGAL**

Projects that produce unprecedented amounts of data, such as the human genome project[15] or the ENCODE project, present new computational and data-analysis challenges and have been a major force driving the development of computational methods in genomics. The human genome project produced one bit of information per DNA base pair, and led to advances in algorithms for sequence matching and alignment. By contrast, in its 1,640 genome-wide data sets, ENCODE provides a profile of the accessibility, methylation, transcriptional status, chromatin structure and bound molecules for every base pair. Processing the project's raw data to obtain this functional information has been an immense effort.

For each of the molecular-profiling methods used, the ENCODE researchers devised novel processing algorithms designed to remove outliers and protocol-specific biases, and to ensure the reliability of the derived functional information. These processing pipelines and quality-control measures have been adapted by the research community as the standard for the analysis of such data. The high quality of the functional information they produce is evident from the exquisite detail and accuracy achieved, such as the ability to observe the crystallographic topography of protein–DNA interfaces in DNase I footprints[5], and the observation of more than one-million-fold variation in dynamic range in the concentrations of different RNA transcripts[3].

> ***"The high quality of the functional information produced is evident from the exquisite detail and accuracy achieved."***

But beyond these individual methods for data processing, the profound biological insights of ENCODE undoubtedly come from computational approaches that integrated multiple data types. For example, by combining data on DNA methylation, DNA accessibility and transcription-factor expression. Thurman *et al.*[4] provide fascinating insight into the causal role of DNA methylation in gene silencing. They find that transcription-factor binding sites are, on average, less frequently methylated in cell types that express those transcription factors, suggesting that binding-site methylation often results from a passive mechanism that methylates sites not bound by transcription factors.

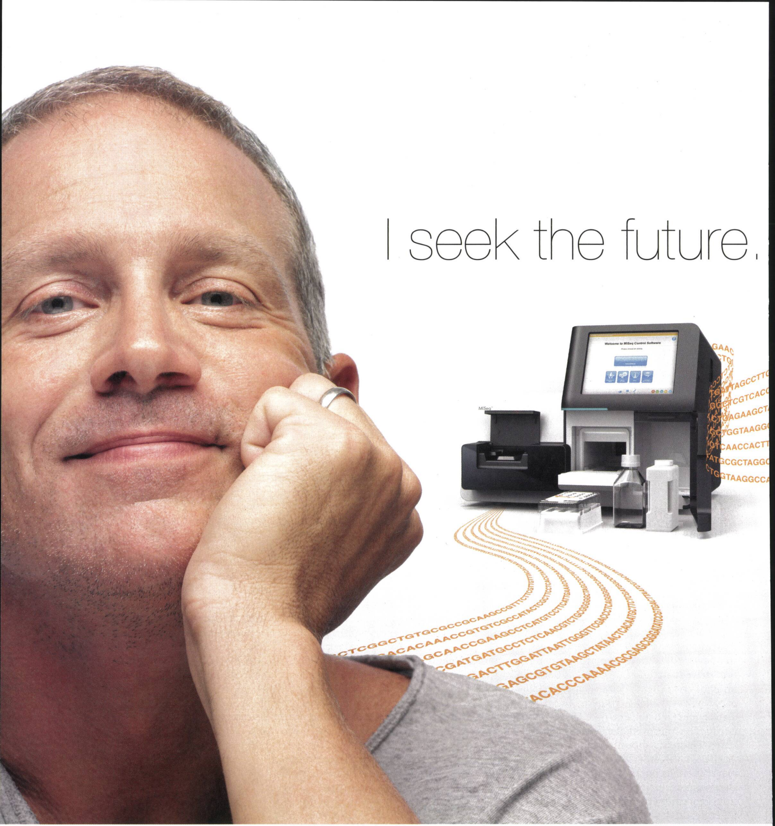
Despite the extensive functional information provided by ENCODE, we are still far from the ultimate goal of understanding the function of the genome in every cell of every person, and across time within the same person. Even if the throughput rate of the ENCODE profiling methods increases dramatically, it is clear that brute-force measurement of this vast space is not feasible. Rather, we must move on from descriptive and correlative computational analyses, and work towards deriving quantitative models that integrate the relevant protein, RNA and chromatin components. We must then describe how these components interact with each other, how they bind the genome and how these binding events regulate transcription.

If successful, such models will be able to predict the genome's function at times and in settings that have not been directly measured. By allowing us to determine which assumptions regarding the physical interactions of the system lead to models that better explain measured patterns, the ENCODE data provide an invaluable opportunity to address this next immense computational challenge. ■

---

**Eran Segal** *is in the Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel.*
*e-mail: eran.segal@weizmann.ac.il*

1. The ENCODE Project Consortium *Science* **306,** 636–640 (2004).
2. The ENCODE Project Consortium *Nature* **489,** 57–74 (2012).
3. Djebali, S. *et al. Nature* **489,** 101–108 (2012).
4. Thurman, R. E. *et al. Nature* **489,** 75–82 (2012).
5. Neph, S. *et al. Nature* **489,** 83–90 (2012).
6. Gerstein, M. B. *et al. Nature* **489,** 91–100 (2012).
7. Sanyal, A., Lajoie, B., Jain, G. & Dekker, J. *Nature* **489,** 109–113 (2012).
8. Birney, E. *et al. Nature* **447,** 799–816 (2007).
9. Mardis, E. R. *Nature* **470,** 198–203 (2011).
10. Gonzaga-Jauregui, C., Lupski, J. R. & Gibbs, R. A. *Annu. Rev. Med.* **63,** 35–61 (2012).
11. Sproul, D. *et al. Proc. Natl Acad. Sci. USA* **108,** 4364–4369 (2011).
12. Carroll, S. B. *Cell* **134,** 25–36 (2008).
13. Prabhakar, S. *et al. Science* **321,** 1346–1350 (2008).
14. Pritchard, J. K., Pickrell, J. K. & Coop, G. *Curr. Biol.* **20,** R208–R215 (2010).
15. Lander, E. S. *et al. Nature* **409,** 860–921 (2001).


I seek the future.
MiSeq
Welcome to MiSeq Control Software

MiSeq. Next-generation sequencing for all you seek.

You want amazing accuracy and performance on your benchtop. Illumina delivers—yet again. The MiSeq Personal Sequencer is the only fully integrated, truly end-to-end benchtop solution around. It's just one more example of why Illumina solutions generate a remarkable 90% of all the world's sequencing data. Discover what's possible.

*www.illumina.com/iseek*

illumina®

doi:10.1038/nature11247

# An integrated encyclopedia of DNA elements in the human genome

The ENCODE Project Consortium*

**The human genome encodes the blueprint of life, but the function of the vast majority of its nearly three billion bases is unknown. The Encyclopedia of DNA Elements (ENCODE) project has systematically mapped regions of transcription, transcription factor association, chromatin structure and histone modification. These data enabled us to assign biochemical functions for 80% of the genome, in particular outside of the well-studied protein-coding regions. Many discovered candidate regulatory elements are physically associated with one another and with expressed genes, providing new insights into the mechanisms of gene regulation. The newly identified elements also show a statistical correspondence to sequence variants linked to human disease, and can thereby guide interpretation of this variation. Overall, the project provides new insights into the organization and regulation of our genes and genome, and is an expansive resource of functional annotations for biomedical research.**

The human genome sequence provides the underlying code for human biology. Despite intensive study, especially in identifying protein-coding genes, our understanding of the genome is far from complete, particularly with regard to non-coding RNAs, alternatively spliced transcripts and regulatory sequences. Systematic analyses of transcripts and regulatory information are essential for the identification of genes and regulatory regions, and are an important resource for the study of human biology and disease. Such analyses can also provide comprehensive views of the organization and variability of genes and regulatory information across cellular contexts, species and individuals.

**ENCODE**
Encyclopedia of DNA Elements
nature.com/encode

The Encyclopedia of DNA Elements (ENCODE) project aims to delineate all functional elements encoded in the human genome[1–3]. Operationally, we define a functional element as a discrete genome segment that encodes a defined product (for example, protein or non-coding RNA) or displays a reproducible biochemical signature (for example, protein binding, or a specific chromatin structure). Comparative genomic studies suggest that 3–8% of bases are under purifying (negative) selection[4–8] and therefore may be functional, although other analyses have suggested much higher estimates[9–11]. In a pilot phase covering 1% of the genome, the ENCODE project annotated 60% of mammalian evolutionarily constrained bases, but also identified many additional putative functional elements without evidence of constraint[2]. The advent of more powerful DNA sequencing technologies now enables whole-genome and more precise analyses with a broad repertoire of functional assays.

Here we describe the production and initial analysis of 1,640 data sets designed to annotate functional elements in the entire human genome. We integrate results from diverse experiments within cell types, related experiments involving 147 different cell types, and all ENCODE data with other resources, such as candidate regions from genome-wide association studies (GWAS) and evolutionarily constrained regions. Together, these efforts reveal important features about the organization and function of the human genome, summarized below.

- The vast majority (80.4%) of the human genome participates in at least one biochemical RNA- and/or chromatin-associated event in at least one cell type. Much of the genome lies close to a regulatory event: 95% of the genome lies within 8 kilobases (kb) of a DNA–protein interaction (as assayed by bound ChIP-seq motifs or DNase I footprints), and 99% is within 1.7 kb of at least one of the biochemical events measured by ENCODE.
- Primate-specific elements as well as elements without detectable mammalian constraint show, in aggregate, evidence of negative selection; thus, some of them are expected to be functional.
- Classifying the genome into seven chromatin states indicates an initial set of 399,124 regions with enhancer-like features and 70,292 regions with promoter-like features, as well as hundreds of thousands of quiescent regions. High-resolution analyses further subdivide the genome into thousands of narrow states with distinct functional properties.
- It is possible to correlate quantitatively RNA sequence production and processing with both chromatin marks and transcription factor binding at promoters, indicating that promoter functionality can explain most of the variation in RNA expression.
- Many non-coding variants in individual genome sequences lie in ENCODE-annotated functional regions; this number is at least as large as those that lie in protein-coding genes.
- Single nucleotide polymorphisms (SNPs) associated with disease by GWAS are enriched within non-coding functional elements, with a majority residing in or near ENCODE-defined regions that are outside of protein-coding genes. In many cases, the disease phenotypes can be associated with a specific cell type or transcription factor.

## ENCODE data production and initial analyses

Since 2007, ENCODE has developed methods and performed a large number of sequence-based studies to map functional elements across the human genome[3]. The elements mapped (and approaches used) include RNA transcribed regions (RNA-seq, CAGE, RNA-PET and manual annotation), protein-coding regions (mass spectrometry), transcription-factor-binding sites (ChIP-seq and DNase-seq), chromatin structure (DNase-seq, FAIRE-seq, histone ChIP-seq and MNase-seq), and DNA methylation sites (RRBS assay) (Box 1 lists methods and abbreviations; Supplementary Table 1, section P, details production statistics)[3]. To compare and integrate results across the different laboratories, data production efforts focused on two selected

---

*Lists of participants and their affiliations appear at the end of the paper.

### BOX 1
### ENCODE abbreviations

**RNA-seq.** Isolation of RNA sequences, often with different purification techniques to isolate different fractions of RNA followed by high-throughput sequencing.

**CAGE.** Capture of the methylated cap at the 5′ end of RNA, followed by high-throughput sequencing of a small tag adjacent to the 5′ methylated caps. 5′ methylated caps are formed at the initiation of transcription, although other mechanisms also methylate 5′ ends of RNA.

**RNA-PET.** Simultaneous capture of RNAs with both a 5′ methyl cap and a poly(A) tail, which is indicative of a full-length RNA. This is then followed by sequencing a short tag from each end by high-throughput sequencing.

**ChIP-seq.** Chromatin immunoprecipitation followed by sequencing. Specific regions of crosslinked chromatin, which is genomic DNA in complex with its bound proteins, are selected by using an antibody to a specific epitope. The enriched sample is then subjected to high-throughput sequencing to determine the regions in the genome most often bound by the protein to which the antibody was directed. Most often used are antibodies to any chromatin-associated epitope, including transcription factors, chromatin binding proteins and specific chemical modifications on histone proteins.

**DNase-seq.** Adaption of established regulatory sequence assay to modern techniques. The DNase I enzyme will preferentially cut live chromatin preparations at sites where nearby there are specific (non-histone) proteins. The resulting cut points are then sequenced using high-throughput sequencing to determine those sites 'hypersensitive' to DNase I, corresponding to open chromatin.

**FAIRE-seq.** Formaldehyde assisted isolation of regulatory elements. FAIRE isolates nucleosome-depleted genomic regions by exploiting the difference in crosslinking efficiency between nucleosomes (high) and sequence-specific regulatory factors (low). FAIRE consists of crosslinking, phenol extraction, and sequencing the DNA fragments in the aqueous phase.

**RRBS.** Reduced representation bisulphite sequencing. Bisulphite treatment of DNA sequence converts unmethylated cytosines to uracil. To focus the assay and save costs, specific restriction enzymes that cut around CpG dinucleotides can reduce the genome to a portion specifically enriched in CpGs. This enriched sample is then sequenced to determine the methylation status of individual cytosines quantitatively.

**Tier 1.** Tier 1 cell types were the highest-priority set and comprised three widely studied cell lines: K562 erythroleukaemia cells; GM12878, a B-lymphoblastoid cell line that is also part of the 1000 Genomes project (http://1000genomes.org)[55]; and the H1 embryonic stem cell (H1 hESC) line.

**Tier 2.** The second-priority set of cell types in the ENCODE project which included HeLa-S3 cervical carcinoma cells, HepG2 hepatoblastoma cells and primary (non-transformed) human umbilical vein endothelial cells (HUVECs).

**Tier 3.** Any other ENCODE cell types not in tier 1 or tier 2.

sets of cell lines, designated 'tier 1' and 'tier 2' (Box 1). To capture a broader spectrum of biological diversity, selected assays were also executed on a third tier comprising more than 100 cell types including primary cells. All data and protocol descriptions are available at http://www.encodeproject.org/, and a User's Guide including details of cell-type choice and limitations was published recently[3].

## Integration methodology

For consistency, data were generated and processed using standardized guidelines, and for some assays, new quality-control measures were designed (see refs 3, 12 and http://encodeproject.org/ENCODE/dataStandards.html; A. Kundaje, personal communication). Uniform data-processing methods were developed for each assay (see Supplementary Information; A. Kundaje, personal communication), and most assay results can be represented both as signal information (a per-base estimate across the genome) and as discrete elements (regions computationally identified as enriched for signal). Extensive processing pipelines were developed to generate each representation (M. M. Hoffman *et al.*, manuscript in preparation and A. Kundaje, personal communication). In addition, we developed the irreproducible discovery rate (IDR)[13] measure to provide a robust and conservative estimate of the threshold where two ranked lists of results from biological replicates no longer agree (that is, are irreproducible), and we applied this to defining sets of discrete elements. We identified, and excluded from most analyses, regions yielding untrustworthy signals likely to be artefactual (for example, multicopy regions). Together, these regions comprise 0.39% of the genome (see Supplementary Information). The poster accompanying this issue represents different ENCODE-identified elements and their genome coverage.

## Transcribed and protein-coding regions

We used manual and automated annotation to produce a comprehensive catalogue of human protein-coding and non-coding RNAs as well as pseudogenes, referred to as the GENCODE reference gene set[14,15] (Supplementary Table 1, section U). This includes 20,687 protein-coding genes (GENCODE annotation, v7) with, on average, 6.3 alternatively spliced transcripts (3.9 different protein-coding transcripts) per locus. In total, GENCODE-annotated exons of protein-coding genes cover 2.94% of the genome or 1.22% for protein-coding exons. Protein-coding genes span 33.45% from the outermost start to stop codons, or 39.54% from promoter to poly(A) site. Analysis of mass spectrometry data from K562 and GM12878 cell lines yielded 57 confidently identified unique peptide sequences in intergenic regions relative to GENCODE annotation. Taken together with evidence of pervasive genome transcription[16], these data indicate that additional protein-coding genes remain to be found.

In addition, we annotated 8,801 automatically derived small RNAs and 9,640 manually curated long non-coding RNA (lncRNA) loci[17]. Comparing lncRNAs to other ENCODE data indicates that lncRNAs are generated through a pathway similar to that for protein-coding genes[17]. The GENCODE project also annotated 11,224 pseudogenes, of which 863 were transcribed and associated with active chromatin[18].

## RNA

We sequenced RNA[16] from different cell lines and multiple subcellular fractions to develop an extensive RNA expression catalogue. Using a conservative threshold to identify regions of RNA activity, 62% of genomic bases are reproducibly represented in sequenced long (>200 nucleotides) RNA molecules or GENCODE exons. Of these bases, only 5.5% are explained by GENCODE exons. Most transcribed bases are within or overlapping annotated gene boundaries (that is, intronic), and only 31% of bases in sequenced transcripts were intergenic[16].

We used CAGE-seq (5′ cap-targeted RNA isolation and sequencing) to identify 62,403 transcription start sites (TSSs) at high confidence (IDR of 0.01) in tier 1 and 2 cell types. Of these, 27,362 (44%) are within 100 base pairs (bp) of the 5′ end of a GENCODE-annotated transcript or previously reported full-length messenger RNA. The remaining regions predominantly lie across exons and 3′ untranslated regions (UTRs), and some exhibit cell-type-restricted expression; these may represent the start sites of novel, cell-type-specific transcripts.

Finally, we saw a significant proportion of coding and non-coding transcripts processed into steady-state stable RNAs shorter than 200 nucleotides. These precursors include transfer RNA, microRNA, small nuclear RNA and small nucleolar RNA (tRNA, miRNA, snRNA and snoRNA, respectively) and the 5′ termini of these processed products align with the capped 5′ end tags[16].

**Table 1 | Summary of transcription factor classes analysed in ENCODE**

| Acronym | Description | Factors analysed |
|---|---|---|
| ChromRem | ATP-dependent chromatin complexes | 5 |
| DNARep | DNA repair | 3 |
| HISase | Histone acetylation, deacetylation or methylation complexes | 8 |
| Other | Cyclin kinase associated with transcription | 1 |
| Pol2 | Pol II subunit | 1 (2 forms) |
| Pol3 | Pol III-associated | 6 |
| TFNS | General Pol II-associated factor, not site-specific | 8 |
| TFSS | Pol II transcription factor with sequence-specific DNA binding | 87 |

## Protein bound regions

To identify regulatory regions directly, we mapped the binding locations of 119 different DNA-binding proteins and a number of RNA polymerase components in 72 cell types using ChIP-seq (Table 1, Supplementary Table 1, section N, and ref. 19); 87 (73%) were sequence-specific transcription factors. Overall, 636,336 binding regions covering 231 megabases (Mb; 8.1%) of the genome are enriched for regions bound by DNA-binding proteins across all cell types. We assessed each protein-binding site for enrichment of known DNA-binding motifs and the presence of novel motifs. Overall, 86% of the DNA segments occupied by sequence-specific transcription factors contained a strong DNA-binding motif, and in most (55%) cases the known motif was most enriched (P. Kheradpour and M. Kellis, manuscript in preparation).

Protein-binding regions lacking high or moderate affinity cognate recognition sites have 21% lower median scores by rank than regions with recognition sequences (Wilcoxon rank sum *P* value $<10^{-16}$). Eighty-two per cent of the low-signal regions have high-affinity recognition sequences for other factors. In addition, when ChIP-seq peaks are ranked by their concordance with their known recognition sequence, the median DNase I accessibility is twofold higher in the bottom 20% of peaks than in the upper 80% (genome structure correction (GSC)[20] *P* value $<10^{-16}$), consistent with previous observations[21–24]. We speculate that low signal regions are either lower-affinity sites[21] or indirect transcription-factor target regions associated through interactions with other factors (see also refs 25, 26).

We organized all the information associated with each transcription factor—including the ChIP-seq peaks, discovered motifs and associated histone modification patterns—in FactorBook (http://www.factorbook.org; ref. 26), a public resource that will be updated as the project proceeds.

## DNase I hypersensitive sites and footprints

Chromatin accessibility characterized by DNase I hypersensitivity is the hallmark of regulatory DNA regions[27,28]. We mapped 2.89 million unique, non-overlapping DNase I hypersensitive sites (DHSs) by DNase-seq in 125 cell types, the overwhelming majority of which lie distal to TSSs[29]. We also mapped 4.8 million sites across 25 cell types that displayed reduced nucleosomal crosslinking by FAIRE, many of which coincide with DHSs. In addition, we used micrococcal nuclease to map nucleosome occupancy in GM12878 and K562 cells[30].

In tier 1 and tier 2 cell types, we identified a mean of 205,109 DHSs per cell type (at false discovery rate (FDR) 1%), encompassing an average of 1.0% of the genomic sequence in each cell type, and 3.9% in aggregate. On average, 98.5% of the occupancy sites of transcription factors mapped by ENCODE ChIP-seq (and, collectively, 94.4% of all 1.1 million transcription factor ChIP-seq peaks in K562 cells) lie within accessible chromatin defined by DNase I hotspots[29]. However, a small number of factors, most prominently heterochromatin-bound repressive complexes (for example, the TRIM28–SETDB1–ZNF274 complex[31,32] encoded by the *TRIM28*, *SETDB1* and *ZNF274* genes), seem to occupy a significant fraction of nucleosomal sites.

Using genomic DNase I footprinting[33,34] on 41 cell types we identified 8.4 million distinct DNase I footprints (FDR 1%)[25]. Our *de novo* motif discovery on DNase I footprints recovered ~90% of known transcription factor motifs, together with hundreds of novel evolutionarily conserved motifs, many displaying highly cell-selective occupancy patterns similar to major developmental and tissue-specific regulators.

## Regions of histone modification

We assayed chromosomal locations for up to 12 histone modifications and variants in 46 cell types, including a complete matrix of eight modifications across tier 1 and tier 2. Because modification states may span multiple nucleosomes, which themselves can vary in position across cell populations, we used a continuous signal measure of histone modifications in downstream analysis, rather than calling regions (M. M. Hoffman *et al.*, manuscript in preparation; see http://code.google.com/p/align2rawsignal/). For the strongest, 'peak-like' histone modifications, we used MACS[35] to characterize enriched sites. Table 2 describes the different histone modifications, their peak characteristics, and a summary of their known roles (reviewed in refs 36–39).

Our data show that global patterns of modification are highly variable across cell types, in accordance with changes in transcriptional activity. Consistent with previous studies[40,41], we find that integration of the different histone modification information can be used systematically to assign functional attributes to genomic regions (see below).

## DNA methylation

Methylation of cytosine, usually at CpG dinucleotides, is involved in epigenetic regulation of gene expression. Promoter methylation is typically associated with repression, whereas genic methylation correlates with transcriptional activity[42]. We used reduced representation bisulphite sequencing (RRBS) to profile DNA methylation quantitatively for an average of 1.2 million CpGs in each of 82 cell lines and tissues (8.6% of non-repetitive genomic CpGs), including CpGs in intergenic regions, proximal promoters and intragenic regions (gene bodies)[43], although it should be noted that the RRBS method preferentially targets CpG-rich islands. We found that 96% of CpGs exhibited differential methylation in at least one cell type or tissue

**Table 2 | Summary of ENCODE histone modifications and variants**

| Histone modification or variant | Signal characteristics | Putative functions |
|---|---|---|
| H2A.Z | Peak | Histone protein variant (H2A.Z) associated with regulatory elements with dynamic chromatin |
| H3K4me1 | Peak/region | Mark of regulatory elements associated with enhancers and other distal elements, but also enriched downstream of transcription starts |
| H3K4me2 | Peak | Mark of regulatory elements associated with promoters and enhancers |
| H3K4me3 | Peak | Mark of regulatory elements primarily associated with promoters/transcription starts |
| H3K9ac | Peak | Mark of active regulatory elements with preference for promoters |
| H3K9me1 | Region | Preference for the 5′ end of genes |
| H3K9me3 | Peak/region | Repressive mark associated with constitutive heterochromatin and repetitive elements |
| H3K27ac | Peak | Mark of active regulatory elements; may distinguish active enhancers and promoters from their inactive counterparts |
| H3K27me3 | Region | Repressive mark established by polycomb complex activity associated with repressive domains and silent developmental genes |
| H3K36me3 | Region | Elongation mark associated with transcribed portions of genes, with preference for 3′ regions after intron 1 |
| H3K79me2 | Region | Transcription-associated mark, with preference for 5′ end of genes |
| H4K20me1 | Region | Preference for 5′ end of genes |

assayed (K. Varley *et al.*, personal communication), and levels of DNA methylation correlated with chromatin accessibility. The most variably methylated CpGs are found more often in gene bodies and intergenic regions, rather than in promoters and upstream regulatory regions. In addition, we identified an unexpected correspondence between unmethylated genic CpG islands and binding by P300, a histone acetyltransferase linked to enhancer activity[44].

Because RRBS is a sequence-based assay with single-base resolution, we were able to identify CpGs with allele-specific methylation consistent with genomic imprinting, and determined that these loci exhibit aberrant methylation in cancer cell lines (K. Varley *et al.*, personal communication). Furthermore, we detected reproducible cytosine methylation outside CpG dinucleotides in adult tissues[45], providing further support that this non-canonical methylation event may have important roles in human biology (K. Varley *et al.*, personal communication).

### Chromosome-interacting regions

Physical interaction between distinct chromosome regions that can be separated by hundreds of kilobases is thought to be important in the regulation of gene expression[46]. We used two complementary chromosome conformation capture (3C)-based technologies to probe these long-range physical interactions.

A 3C-carbon copy (5C) approach[47,48] provided unbiased detection of long-range interactions with TSSs in a targeted 1% of the genome (the 44 ENCODE pilot regions) in four cell types (GM12878, K562, HeLa-S3 and H1 hESC)[49]. We discovered hundreds of statistically significant long-range interactions in each cell type after accounting for chromatin polymer behaviour and experimental variation. Pairs of interacting loci showed strong correlation between the gene expression level of the TSS and the presence of specific functional element classes such as enhancers. The average number of distal elements interacting with a TSS was 3.9, and the average number of TSSs interacting with a distal element was 2.5, indicating a complex network of interconnected chromatin. Such interwoven long-range architecture was also uncovered genome-wide using chromatin interaction analysis with paired-end tag sequencing (ChIA-PET)[50] applied to identify interactions in chromatin enriched by RNA polymerase II (Pol II) ChIP from five cell types[51]. In K562 cells, we identified 127,417 promoter-centred chromatin interactions using ChIA-PET, 98% of which were intra-chromosomal. Whereas promoter regions of 2,324 genes were involved in 'single-gene' enhancer–promoter interactions, those of 19,813 genes were involved in 'multi-gene' interaction complexes spanning up to several megabases, including promoter–promoter and enhancer–promoter interactions[51].

These analyses portray a complex landscape of long-range gene–element connectivity across ranges of hundreds of kilobases to several megabases, including interactions among unrelated genes (Supplementary Fig. 1, section Y). Furthermore, in the 5C results, 50–60% of long-range interactions occurred in only one of the four cell lines, indicative of a high degree of tissue specificity for gene–element connectivity[49].

### Summary of ENCODE-identified elements

Accounting for all these elements, a surprisingly large amount of the human genome, 80.4%, is covered by at least one ENCODE-identified element (detailed in Supplementary Table 1, section Q). The broadest element class represents the different RNA types, covering 62% of the genome (although the majority is inside of introns or near genes). Regions highly enriched for histone modifications form the next largest class (56.1%). Excluding RNA elements and broad histone elements, 44.2% of the genome is covered. Smaller proportions of the genome are occupied by regions of open chromatin (15.2%) or sites of transcription factor binding (8.1%), with 19.4% covered by at least one DHS or transcription factor ChIP-seq peak across all cell lines. Using our most conservative assessment, 8.5% of bases are covered by either a transcription-factor-binding-site motif (4.6%) or a DHS footprint (5.7%). This, however, is still about 4.5-fold higher than the amount of protein-coding exons, and about twofold higher than the estimated amount of pan-mammalian constraint.

Given that the ENCODE project did not assay all cell types, or all transcription factors, and in particular has sampled few specialized or developmentally restricted cell lineages, these proportions must be underestimates of the total amount of functional bases. However, many assays were performed on more than one cell type, allowing assessment of the rate of discovery of new elements. For both DHSs and CTCF-bound sites, the number of new elements initially increases rapidly with a steep gradient for the saturation curve and then slows with increasing number of cell types (Supplementary Figs 1 and 2, section R). With the current data, at the flattest part of the saturation curve each new cell type adds, on average, 9,500 DHS elements (across 106 cell types) and 500 CTCF-binding elements (across 49 cell types), representing 0.45% of the total element number. We modelled saturation for the DHSs and CTCF-binding sites using a Weibull distribution ($r^2 > 0.999$) and predict saturation at approximately 4.1 million (standard error (s.e.) = 108,000) and 185,100 (s.e. = 18,020) sites, respectively, indicating that we have discovered around half of the estimated total DHSs. These estimates represent a lower bound, but reinforce the observation that there is more non-coding functional DNA than either coding sequence or mammalian evolutionarily constrained bases.

### The impact of selection on functional elements

From comparative genomic studies, at least 3–8% of bases are under purifying (negative) selection[4–11], indicating that these bases may potentially be functional. We previously found that 60% of mammalian evolutionarily constrained bases were annotated in the ENCODE pilot project, but also observed that many functional elements lacked evidence of constraint[2], a conclusion substantiated by others[52–54]. The diversity and genome-wide occurrence of functional elements now identified provides an unprecedented opportunity to examine further the forces of negative selection on human functional sequences.

We examined negative selection using two measures that highlight different periods of selection in the human genome. The first measure, inter-species, pan-mammalian constraint (GERP-based scores; 24 mammals[8]), addresses selection during mammalian evolution. The second measure is intra-species constraint estimated from the numbers of variants discovered in human populations using data from the 1000 Genomes project[55], and covers selection over human evolution. In Fig. 1, we plot both these measures of constraint for different classes of identified functional elements, excluding features overlapping exons and promoters that are known to be constrained. Each graph also shows genomic background levels and measures of coding-gene constraint for comparison. Because we plot human population diversity on an inverted scale, elements that are more constrained by negative selection will tend to lie in the upper and right-hand regions of the plot.

For DNase I elements (Fig. 1b) and bound motifs (Fig. 1c), most sets of elements show enrichment in pan-mammalian constraint and decreased human population diversity, although for some cell types the DNase I sites do not seem overall to be subject to pan-mammalian constraint. Bound transcription factor motifs have a natural control from the set of transcription factor motifs with equal sequence potential for binding but without binding evidence from ChIP-seq experiments—in all cases, the bound motifs show both more mammalian constraint and higher suppression of human diversity.

Consistent with previous findings, we do not observe genome-wide evidence for pan-mammalian selection of novel RNA sequences (Fig. 1d). There are also a large number of elements without mammalian constraint, between 17% and 90% for transcription-factor-binding regions as well as DHSs and FAIRE regions. Previous studies could not determine whether these sequences are either biochemically active, but with little overall impact on the organism, or under lineage-specific selection. By isolating sequences preferentially inserted into

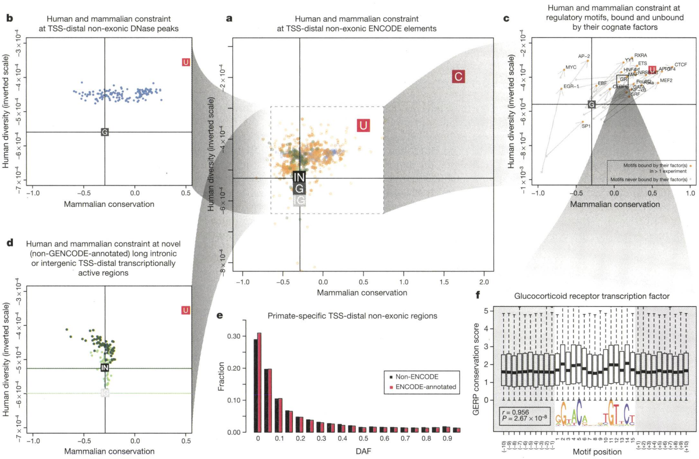


**Figure 1 | Impact of selection on ENCODE functional elements in mammals and human populations.** **a**, Levels of pan-mammalian constraint (mean GERP score; 24 mammals[8], *x* axis) compared to diversity, a measure of negative selection in the human population (mean expected heterozygosity, inverted scale, *y* axis) for ENCODE data sets. Each point is an average for a single data set. The top-right corners have the strongest evolutionary constraint and lowest diversity. Coding (C), UTR (U), genomic (G), intergenic (IG) and intronic (IN) averages are shown as filled squares. In each case the vertical and horizontal cross hairs show representative levels for the neutral expectation for mammalian conservation and human population diversity, respectively. The spread over all non-exonic ENCODE elements greater than 2.5 kb from TSSs is shown. The inner dashed box indicates that parts of the plot have been magnified for the surrounding outer panels, although the scales in the outer plots provide the exact regions and dimensions magnified. The spread for DHS sites (**b**) and RNA elements (**d**) is shown in the plots on the left. RNA elements are either long novel intronic (dark green) or long intergenic (light green) RNAs. The horizontal cross hairs are colour-coded to the relevant data set in **d**. **c**, Spread of transcription factor motif instances either in regions bound by the transcription factor (orange points) or in the corresponding unbound motif matches in grey, with bound and unbound points connected with an arrow in each case showing that bound sites are generally more constrained and less diverse. **e**, Derived allele frequency spectrum for primate-specific elements, with variations outside ENCODE elements in black and variations covered by ENCODE elements in red. The increase in low-frequency alleles compared to background is indicative of negative selection occurring in the set of variants annotated by the ENCODE data. **f**, Aggregation of mammalian constraint scores over the glucocorticoid receptor (GR) transcription factor motif in bound sites, showing the expected correlation with the information content of bases in the motif. An interactive version of this figure is available in the online version of the paper.

the primate lineage, which is only feasible given the genome-wide scale of this data, we are able to examine this issue specifically. Most primate-specific sequence is due to retrotransposon activity, but an appreciable proportion is non-repetitive primate-specific sequence. Of 104,343,413 primate-specific bases (excluding repetitive elements), 67,769,372 (65%) are found within ENCODE-identified elements. Examination of 227,688 variants segregating in these primate-specific regions revealed that all classes of elements (RNA and regulatory) show depressed derived allele frequencies, consistent with recent negative selection occurring in at least some of these regions (Fig. 1e). An alternative approach examining sequences that are not clearly under pan-mammalian constraint showed a similar result (L. Ward and M. Kellis, manuscript submitted). This indicates that an appreciable proportion of the unconstrained elements are lineage-specific elements required for organismal function, consistent with long-standing views of recent evolution[56], and the remainder are probably 'neutral' elements[2] that are not currently under selection but may still affect cellular or larger scale phenotypes without an effect on fitness.

The binding patterns of transcription factors are not uniform, and we can correlate both inter- and intra-species measures of negative selection with the overall information content of motif positions. The selection on some motif positions is as high as protein-coding exons (Fig. 1f; L. Ward and M. Kellis, manuscript submitted). These aggregate measures across motifs show that the binding preferences found in the population of sites are also relevant to the per-site behaviour. By developing a per-site metric of population effect on bound motifs, we found that highly constrained bound instances across mammals are able to buffer the impact of individual variation[57].

## ENCODE data integration with known genomic features

### Promoter-anchored integration

Many of the ENCODE assays directly or indirectly provide information about the action of promoters. Focusing on the TSSs of protein-coding transcripts, we investigated the relationships between different ENCODE assays, in particular testing the hypothesis that RNA expression (output) can be effectively predicted from patterns of chromatin modification or transcription factor binding (input). Consistent with previous reports[58], we observe two relatively distinct types of promoter: (1) broad, mainly (C+G)-rich, TATA-less promoters; and (2) narrow, TATA-box-containing promoters. These promoters have distinct patterns of histone modifications, and transcription-factor-binding sites are selectively enriched in each class (Supplementary Fig. 1, section Z).

We developed predictive models to explore the interaction between histone modifications and measures of transcription at promoters, distinguishing between modifications known to be added as a consequence of transcription (such as H3K36me3 and H3K79me2) and other categories of histone marks[59]. In our analyses, the best models had two components: an initial classification component (on/off) and a second quantitative model component. Our models showed that activating acetylation marks (H3K27ac and H3K9ac) are roughly as informative as activating methylation marks (H3K4me3 and H3K4me2) (Fig. 2a). Although repressive marks, such as H3K27me3 or H3K9me3, show negative correlation both individually and in the model, removing these marks produces only a small reduction in model performance. However, for a subset of promoters in each cell line, repressive histone marks (H3K27me3 or H3K9me3) must be used to predict their expression accurately. We also examined the interplay between the H3K79me2 and H3K36me3 marks, both of which mark gene bodies, probably reflecting recruitment of modification enzymes by polymerase isoforms. As described previously, H3K79me2 occurs preferentially at the 5′ ends of gene bodies and H3K36me3 occurs more 3′, and our analyses support the previous model in which the H3K79me2 to H3K36me3 transition occurs at the first 3′ splice site[60].

Few previous studies have attempted to build qualitative or quantitative models of transcription genome-wide from transcription factor levels because of the paucity of documented transcription-factor-binding regions and the lack of coordination around a single cell line. We thus examined the predictive capacity of transcription-factor-binding signals for the expression levels of promoters (Fig. 2b).

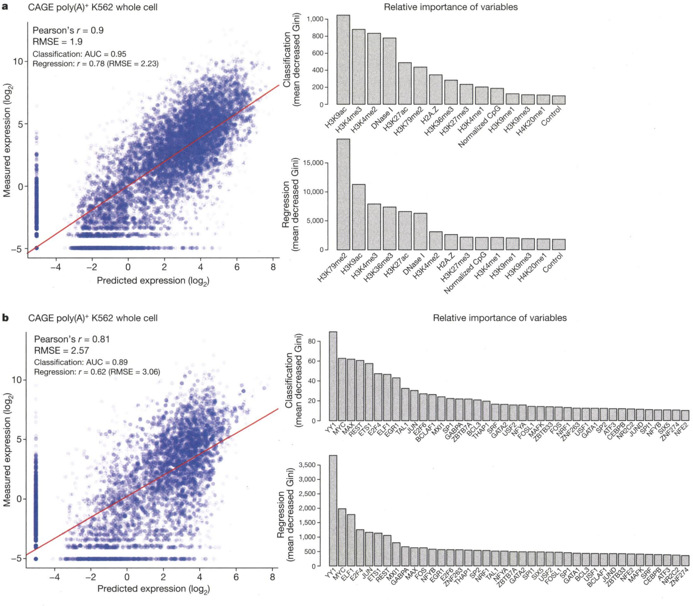


**Figure 2 | Modelling transcription levels from histone modification and transcription-factor-binding patterns. a, b,** Correlative models between either histone modifications or transcription factors, respectively, and RNA production as measured by CAGE tag density at TSSs in K562 cells. In each case the scatter plot shows the output of the correlation models (*x* axis) compared to observed values (*y* axis). The bar graphs show the most important histone modifications (**a**) or transcription factors (**b**) in both the initial classification phase (top bar graph) or the quantitative regression phase (bottom bar graph), with larger values indicating increasing importance of the variable in the model. Further analysis of other cell lines and RNA measurement types is reported elsewhere[59,79]. AUC, area under curve; Gini, Gini coefficient; RMSE, root mean square error.

In contrast to the profiles of histone modifications, most transcription factors show enriched binding signals in a narrow DNA region near the TSS, with relatively higher binding signals in promoters with higher CpG content. Most of this correlation could be recapitulated by looking at the aggregate binding of transcription factors without specific transcription factor terms. Together, these correlation models indicate both that a limited set of chromatin marks are sufficient to 'explain' transcription and that a variety of transcription factors might have broad roles in general transcription levels across many genes. It is important to note that this is an inherently observational study of correlation patterns, and is consistent with a variety of mechanistic models with different causal links between the chromatin, transcription factor and RNA assays. However, it does indicate that there is enough information present at the promoter regions of genes to explain most of the variation in RNA expression.

We developed predictive models similar to those used to model transcriptional activity to explore the relationship between levels of histone modification and inclusion of exons in alternately spliced transcripts. Even accounting for expression level, H3K36me3 has a positive contribution to exon inclusion, whereas H3K79me2 has a negative contribution (H. Tilgner *et al.*, manuscript in preparation). By monitoring the RNA populations in the subcellular fractions of K562 cells, we found that essentially all splicing is co-transcriptional[61], further supporting a link between chromatin structure and splicing.

## Transcription-factor-binding site-anchored integration

Transcription-factor-binding sites provide a natural focus around which to explore chromatin properties. Transcription factors are often multifunctional and can bind a variety of genomic loci with different combinations and patterns of chromatin marks and nucleosome organization. Hence, rather than averaging chromatin mark profiles across all binding sites of a transcription factor, we developed a clustering procedure, termed the Clustered Aggregation Tool (CAGT), to identify subsets of binding sites sharing similar but distinct patterns of chromatin mark signal magnitude, shape and hidden directionality[30]. For example, the average profile of the repressive histone mark H3K27me3 over all 55,782 CTCF-binding sites in H1 hESCs shows poor signal enrichment (Fig. 3a). However, after grouping profiles by signal magnitude we found a subset of 9,840 (17.6%) CTCF-binding sites that exhibit significant flanking H3K27me3 signal. Shape and orientation analysis further revealed that the predominant signal profile for H3K27me3 around CTCF peak summits is asymmetric, consistent with a boundary role for some CTCF sites between active and polycomb-silenced domains. Further examples are provided in Supplementary Figs 5 and 6 of section E. For TAF1, predominantly found near TSSs, the asymmetric sites are orientated with the direction of transcription. However, for distal sites, such as those bound by GATA1 and CTCF, we also observed a high proportion of asymmetric histone patterns, although independent of motif directionality. In fact, all transcription-factor-binding data sets in all cell lines show predominantly asymmetric patterns (asymmetry ratio >0.6) for all chromatin marks but not for DNase I signal (Fig. 3b). This indicates that most transcription-factor-bound chromatin events correlate with structured, directional patterns of histone modifications, and that promoter directionality is not the only source of orientation at these sites.

We also examined nucleosome occupancy relative to the symmetry properties of chromatin marks around transcription-factor-binding sites. Around TSSs, there is usually strong asymmetric nucleosome occupancy, often accounting for most of the histone modification signal (for instance, see Supplementary Fig. 4, section E). However, away from TSSs, there is far less concordance. For example, CTCF-binding sites typically show arrays of well-positioned nucleosomes on either side of the peak summit (Supplementary Fig. 1, section E)[62]. Where the flanking chromatin mark signal is high, the signals are often asymmetric, indicating differential marking with histone modifications (Supplementary Figs 2 and 3, section E). Thus, we confirm on a genome-wide scale that transcription factors can form barriers around which nucleosomes and histone modifications are arranged in a variety of configurations[62–65]. This is explored in further detail in refs 25, 26 and 30.

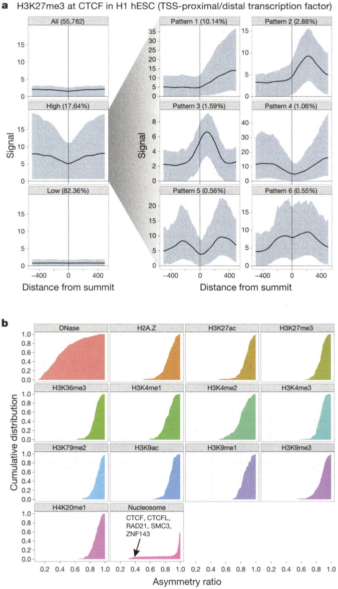


**Figure 3 | Patterns and asymmetry of chromatin modification at transcription-factor-binding sites.** **a**, Results of clustered aggregation of H3K27me3 modification signal around CTCF-binding sites (a multifunctional protein involved with chromatin structure). The first three plots (left column) show the signal behaviour of the histone modification over all sites (top) and then split into the high and low signal components. The solid lines show the mean signal distribution by relative position with the blue shaded area delimiting the tenth and ninetieth percentile range. The high signal component is then decomposed further into six different shape classes on the right (see ref. 30 for details). The shape decomposition process is strand aware. **b**, Summary of shape asymmetry for DNase I, nucleosome and histone modification signals by plotting an asymmetry ratio for each signal over all transcription-factor-binding sites. All histone modifications measured in this study show predominantly asymmetric patterns at transcription-factor-binding sites. An interactive version of this figure is available in the online version of the paper.

## Transcription factor co-associations

Transcription-factor-binding regions are nonrandomly distributed across the genome, with respect to both other features (for example, promoters) and other transcription-factor-binding regions. Within the

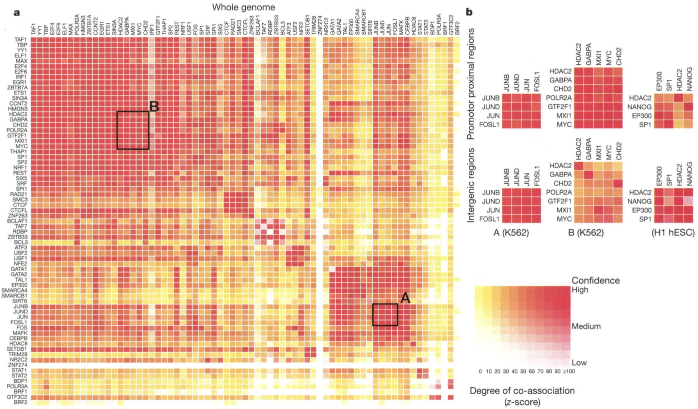


**Figure 4 | Co-association between transcription factors.** **a**, Significant co-associations of transcription factor pairs using the GSC statistic across the entire genome in K562 cells. The colour strength represents the extent of association (from red (strongest), orange, to yellow (weakest)), whereas the depth of colour represents the fit to the GSC[20] model (where white indicates that the statistical model is not appropriate) as indicated by the key. Most transcription factors have a nonrandom association to other transcription factors, and these associations are dependent on the genomic context, meaning that once the genome is separated into promoter proximal and distal regions, the overall levels of co-association decrease, but more specific relationships are uncovered. **b**, Three classes of behaviour are shown. The first column shows a set of associations for which strength is independent of location in promoter and distal regions, whereas the second column shows a set of transcription factors that have stronger associations in promoter-proximal regions. Both of these examples are from data in K562 cells and are highlighted on the genome-wide co-association matrix (**a**) by the labelled boxes A and B, respectively. The third column shows a set of transcription factors that show stronger association in distal regions (in the H1 hESC line). An interactive version of this figure is available in the online version of the paper.

tier 1 and 2 cell lines, we found 3,307 pairs of statistically co-associated factors ($P < 1 \times 10^{-16}$, GSC) involving 114 out of a possible 117 factors (97%) (Fig. 4a). These include expected associations, such as Jun and Fos, and some less expected novel associations, such as TCF7L2 with HNF4-α and FOXA2 (ref. 66; a full listing is given in Supplementary Table 1, section F). When one considers promoter and intergenic

**Table 3 | Summary of the combined state types**

| Label | Description | Details* | Colour |
|---|---|---|---|
| CTCF | CTCF-enriched element | Sites of CTCF signal lacking histone modifications, often associated with open chromatin. Many probably have a function in insulator assays, but because of the multifunctional nature of CTCF, we are conservative in our description. Also enriched for the cohesin components RAD21 and SMC3; CTCF is known to recruit the cohesin complex. | Turquoise |
| E | Predicted enhancer | Regions of open chromatin associated with H3K4me1 signal. Enriched for other enhancer-associated marks, including transcription factors known to act at enhancers. In enhancer assays, many of these (>50%) function as enhancers. A more conservative alternative would be *cis*-regulatory regions. Enriched for sites for the proteins encoded by *EP300*, *FOS*, *FOSL1*, *GATA2*, *HDAC8*, *JUNB*, *JUND*, *NFE2*, *SMARCA4*, *SMARCB1*, *SIRT6* and *TAL1* genes in K562 cells. Have nuclear and whole-cell RNA signal, particularly poly(A)− fraction. | Orange |
| PF | Predicted promoter flanking region | Regions that generally surround TSS segments (see below). | Light red |
| R | Predicted repressed or low-activity region | This is a merged state that includes H3K27me3 polycomb-enriched regions, along with regions that are silent in terms of observed signal for the input assays to the segmentations (low or no signal). They may have other signals (for example, RNA, not in the segmentation input data). Enriched for sites for the proteins encoded by *REST* and some other factors (for example, proteins encoded by *BRF2*, *CEBPB*, *MAFK*, *TRIM28*, *ZNF274* and *SETDB1* genes in K562 cells). | Grey |
| TSS | Predicted promoter region including TSS | Found close to or overlapping GENCODE TSS sites. High precision/recall for TSSs. Enriched for H3K4me3. Sites of open chromatin. Enriched for transcription factors known to act close to promoters and polymerases Pol II and Pol III. Short RNAs are most enriched in these segments. | Bright red |
| T | Predicted transcribed region | Overlap gene bodies with H3K36me3 transcriptional elongation signal. Enriched for phosphorylated form of Pol II signal (elongating polymerase) and poly(A)$^+$ RNA, especially cytoplasmic. | Dark green |
| WE | Predicted weak enhancer or open chromatin *cis*-regulatory element | Similar to the E state, but weaker signals and weaker enrichments. | Yellow |

* Where specific enrichments or overlaps are identified, these are derived from analysis in GM12878 and/or K562 cells where the data for comparison is richest. The colours indicated are used in Figs 5 and 7 and in display of these tracks from the ENCODE data hub.

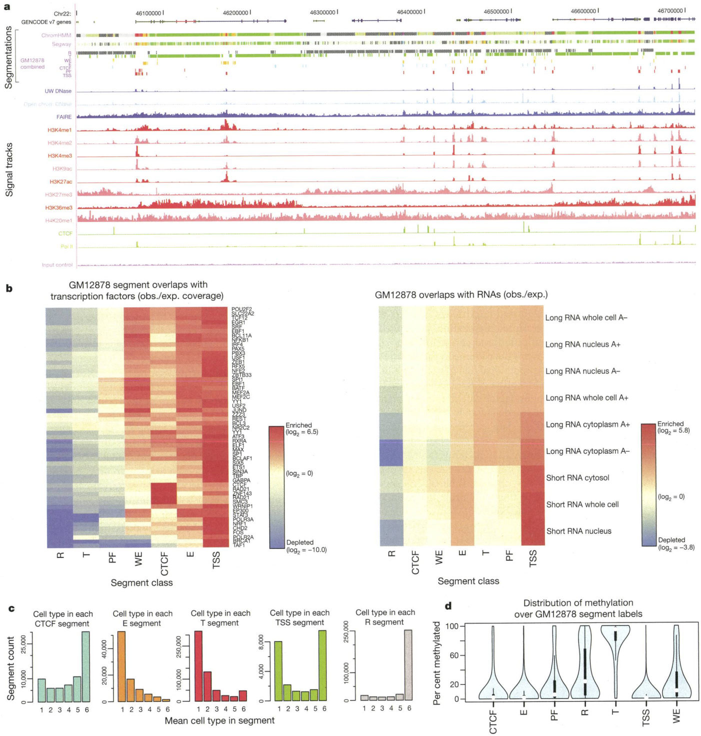


**Figure 5 | Integration of ENCODE data by genome-wide segmentation.** **a**, Illustrative region with the two segmentation methods (ChromHMM and Segway) in a dense view and the combined segmentation expanded to show each state in GM12878 cells, beneath a compressed view of the GENCODE gene annotations. Note that at this level of zoom and genome browser resolution, some segments appear to overlap although they do not. Segmentation classes are named and coloured according to the scheme in Table 3. Beneath the segmentations are shown each of the normalized signals that were used as the input data for the segmentations. Open chromatin signals from DNase-seq from the University of Washington group (UW DNase) or the ENCODE open chromatin group (Openchrom DNase) and FAIRE assays are shown in blue; signal from histone modification ChIP-seq in red; and transcription factor ChIP-seq signal for Pol II and CTCF in green. The mauve ChIP-seq control signal (input control) at the bottom was also included as an input to the segmentation. **b**, Association of selected transcription factor (left) and RNA (right) elements in the combined segmentation states (*x* axis) expressed as an observed/expected ratio (obs./exp.) for each combination of transcription factor or RNA element and segmentation class using the heat-map scale shown in the key besides each heat map. **c**, Variability of states between cell lines, showing the distribution of occurrences of the state in the six cell lines at specific genome locations: from unique to one cell line to ubiquitous in all six cell lines for five states (CTCF, E, T, TSS and R). **d**, Distribution of methylation level at individual sites from RRBS analysis in GM12878 cells across the different states, showing the expected hypomethylation at TSSs and hypermethylation of genes bodies (T state) and repressed (R) regions.

regions separately, this changes to 3,201 pairs (116 factors, 99%) for promoters and 1,564 pairs (108 factors, 92%) for intergenic regions, with some associations more specific to these genomic contexts (for example, the cluster of HDAC2, GABPA, CHD2, GTF2F1, MXI1 and MYC in promoter regions and SP1, EP300, HDAC2 and NANOG in intergenic regions (Fig. 4b)). These general and context-dependent associations lead to a network representation of the co-binding with many interesting properties, explored in refs 19, 25 and 26. In addition, we also identified a set of regions bound by multiple factors representing high occupancy of transcription factor (HOT) regions[67].

## Genome-wide integration

To identify functional regions genome-wide, we next integrated elements independent of genomic landmarks using either discriminative training methods, where a subset of known elements of a particular class were used to train a model that was then used to discover more instances of this class, or using methods in which only data from ENCODE assays were used without explicit knowledge of any annotation.

For discriminative training, we used a three-step process to predict potential enhancers, described in Supplementary Information and ref. 67. Two alternative discriminative models converged on a set of ~13,000 putative enhancers in K562 cells[67]. In the second approach, two methodologically distinct unbiased approaches (see refs 40, 68 and M. M. Hoffman *et al.*, manuscript in preparation) converged on a concordant set of histone modification and chromatin-accessibility patterns that can be used to segment the genome in each of the tier 1 and tier 2 cell lines, although the individual loci in each state in each cell line are different. With the exception of RNA polymerase II and CTCF, the addition of transcription factor data did not substantially alter these patterns. At this stage, we deliberately excluded RNA and methylation assays, reserving these data as a means to validate the segmentations.

Our integration of the two segmentation methods (M. M. Hoffman *et al.*, manuscript in preparation) established a consensus set of seven major classes of genome states, described in Table 3. The standard view of active promoters, with a distinct core promoter region (TSS and PF states), leading to active gene bodies (T, transcribed state), is rediscovered in this model (Fig. 5a, b). There are three 'active' distal states. We tentatively labelled two as enhancers (predicted enhancers, E, and predicted weak enhancers, WE) due to their occurrence in regions of open chromatin with high H3K4me1, although they differ in the levels of marks such as H3K27ac, currently thought to distinguish active from inactive enhancers. The other active state (CTCF) has high CTCF binding and includes sequences that function as insulators in a transfection assay. The remaining repressed state (R) summarizes sequences split between different classes of actively repressed or inactive, quiescent chromatin. We found that the CTCF-binding-associated state is relatively invariant across cell types, with individual regions frequently occupying the CTCF state across all six cell types (Fig. 5c). Conversely, the E and T states have substantial cell-specific behaviour, whereas the TSS state has a bimodal behaviour with similar numbers of cell-invariant and cell-specific occurrences. It is important to note that the consensus summary classes do not capture all the detail discovered in the individual segmentations containing more states.

The distribution of RNA species across segments is quite distinct, indicating that underlying biological activities are captured in the segmentation. Polyadenylated RNA is heavily enriched in gene bodies. Around promoters, there are short RNA species previously identified as promoter-associated short RNAs (Fig. 5b)[16,69]. Similarly, DNA methylation shows marked distinctions between segments, recapitulating the known biology of predominantly unmethylated active promoters (TSS states) followed by methylated gene bodies[42] (T state, Fig. 5d). The two enhancer-enriched states show distinct patterns of DNA methylation, with the less active enhancer state (by H3K27ac/H3K4me1 levels) showing higher methylation. These states also have an excess of RNA elements without poly(A) tails and methyl-cap RNA, as assayed by CAGE sequences, compared to matched intergenic controls, indicating a specific transcriptional mode associated with active enhancers[70]. Transcription factors also showed distinct distributions across the segments (Fig. 5b). A striking pattern is the concentration of transcription factors in the TSS-associated state. The enhancers contain a different set of transcription factors. For example, in K562 cells, the E state is enriched for binding by the proteins encoded by the *EP300*, *FOS*, *FOSL1*, *GATA2*, *HDAC8*, *JUNB*, *JUND*, *NFE2*, *SMARCA4*, *SMARCB1*, *SIRT6* and *TAL1* genes. We tested a subset of these predicted enhancers in both mouse and fish transgenic models (examples in Fig. 6), with over half of the elements showing activity, often in the corresponding tissue type.

The segmentation provides a linear determination of functional state across the genome, but not an association of particular distal regions with genes. By using the variation of DNase I signal across cell lines, 39% of E (enhancer associated) states could be linked to a proposed regulated gene[29] concordant with physical proximity patterns determined by 5C[49] or ChIA-PET.

To provide a fine-grained regional classification, we turned to a self organizing map (SOM) to cluster genome segmentation regions based on their assay signal characteristics (Fig. 7). The segmentation regions were initially randomly assigned to a 1,350-state map in a two-dimensional toroidal space (Fig. 7a). This map can be visualized as a two-dimensional rectangular plane onto which the various signal distributions can be plotted. For instance, the rectangle at the bottom left of Fig. 7a shows the distribution of the genome in the initial randomized map. The SOM was then trained using the twelve different ChIP-seq and DNase-seq assays in the six cell types previously analysed in the large-scale segmentations (that is, over 72-dimensional space). After training, the SOM clustering was again visualized in two dimensions, now showing the organized distribution of genome segments (lower right of panel, Fig. 7a). Individual data sets associated with the genome segments in each SOM map unit (hexagonal cells) can then be visualized in the same framework to learn how each additional kind of data is distributed on the chromatin state map. Figure 7b shows CAGE/TSS expression data overlaid on the randomly initialized (left) and trained map (right) panels. In this way the trained SOM highlighted cell-type-specific TSS clusters (bottom panels of Fig. 7b), indicating that there are sets of tissue-specific TSSs that are distinguished from each other by subtle combinations of ENCODE

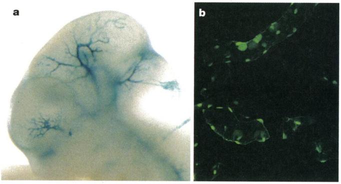


**Figure 6 | Experimental characterization of segmentations.** Randomly sampled E state segments (see Table 3) from the K562 segmentation were cloned for mouse- and fish-based transgenic enhancer assays. **a**, Representative LacZ-stained transgenic embryonic day (E)11.5 mouse embryo obtained with construct hs2065 (EN167, chr10: 46052882–46055670, GRCh37). Highly reproducible staining in the blood vessels was observed in 9 out of 9 embryos resulting from independent transgenic integration events. **b**, Representative green fluorescent protein reporter transgenic medaka fish obtained from a construct with a basal *hsp70* promoter on meganuclease-based transfection. Reproducible transgenic expression in the circulating nucleated blood cells and the endothelial cell walls was seen in 81 out of 100 transgenic tests of this construct.

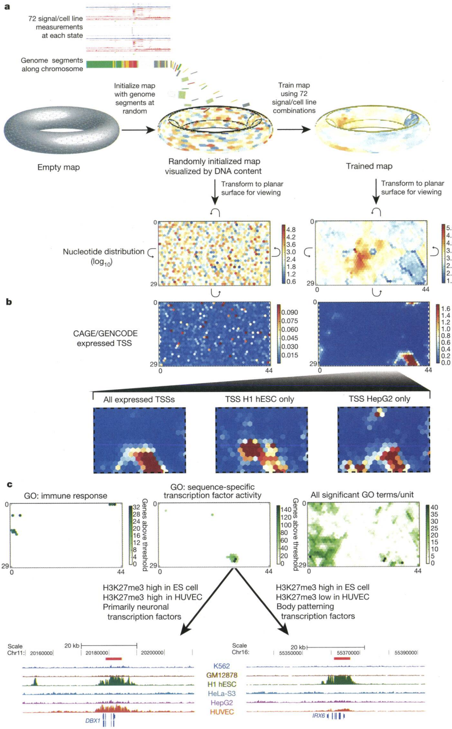


**Figure 7 | High-resolution segmentation of ENCODE data by self-organizing maps (SOM).** **a–c**, The training of the SOM (**a**) and analysis of the results (**b**, **c**) are shown. Initially we arbitrarily placed genomic segments from the ChromHMM segmentation on to the toroidal map surface, although the SOM does not use the ChromHMM state assignments (**a**). We then trained the map using the signal of the 12 different ChIP-seq and DNase-seq assays in the six cell types analysed. Each unit of the SOM is represented here by a hexagonal cell in a planar two-dimensional view of the toroidal map. Curved arrows indicate that traversing the edges of two dimensional view leads back to the opposite edge. The resulting map can be overlaid with any class of ENCODE or other data to view the distribution of that data within this high-resolution segmentation. In panel **a** the distributions of genome bases across the untrained and trained map (left and right, respectively) are shown using heat-map colours for $\log_{10}$ values. **b**, The distribution of TSSs from CAGE experiments of GENCODE annotation on the planar representations of either the initial random organization (left) or the final trained SOM (right) using heat maps coloured according to the accompanying scales. The bottom half of **b** expands the different distributions in the SOM for all expressed TSSs (left) or TSSs specifically expressed in two example cell lines, H1 hESC (centre) and HepG2 (right). **c**, The association of Gene Ontology (GO) terms on the same representation of the same trained SOM. We assigned genes that are within 20 kb of a genomic segment in a SOM unit to that unit, and then associated this set of genes with GO terms using a hypergeometric distribution after correcting for multiple testing. Map units that are significantly associated to GO terms are coloured green, with increasing strength of colour reflecting increasing numbers of genes significantly associated with the GO terms for either immune response (left) or sequence-specific transcription factor activity (centre). In each case, specific SOM units show association with these terms. The right-hand panel shows the distribution on the same SOM of all significantly associated GO terms, now colouring by GO term count per SOM unit. For sequence-specific transcription factor activity, two example genomic regions are extracted at the bottom of panel **c** from neighbouring SOM units. These are regions around the *DBX1* (from SOM unit 26,31, left panel) and *IRX6* (SOM unit 27,30, right panel) genes, respectively, along with their H3K27me3 ChIP-seq signal for each of the tier 1 and 2 cell types. For *DBX1*, representative of a set of primarily neuronal transcription factors associated with unit 26,31, there is a repressive H3K27me3 signal in both H1 hESCs and HUVECs; for *IRX6*, representative of a set of body patterning transcription factors associated with SOM unit 27,30, the repressive mark is restricted largely to the embryonic stem (ES) cell. An interactive version of this figure is available in the online version of the paper.

chromatin data. Many of the ultra-fine-grained state classifications revealed in the SOM are associated with specific gene ontology (GO) terms (right panel of Fig. 7c). For instance, the left panel of Fig. 7c identifies ten SOM map units enriched with genomic regions associated with genes associated with the GO term 'immune response'. The central panel identifies a different set of map units enriched for the GO term 'sequence-specific transcription factor activity'. The two map units most enriched for this GO term, indicated by the darkest green colouring, contain genes with segments that are high in H3K27me3 in H1 hESCs, but that differ in H3K27me3 levels in HUVECs. Gene function analysis with the GO ontology tool (GREAT[71]) reveals that the map unit with high H3K27me3 levels in both cell types is enriched in transcription factor genes with known neuronal functions, whereas the neighbouring map unit is enriched in genes involved in body patterning. The genome browser shots at the bottom of Fig. 7c pick out an example region for each of the two SOM map units illustrating the difference in H3K27me3 signal. Overall, we have 228 distinct GO terms associated with specific segments across one or more states (A. Mortazavi, personal communication), and can assign over one-third of genes to a GO annotation solely on the basis of its multicellular histone patterns. Thus, the SOM analysis provides a fine-grained map of chromatin data across multiple cell types, which can then be used to relate chromatin structure to other data types at differing levels of resolution (for instance, the large cluster of units containing any active TSS, its subclusters composed of units enriched in TSSs active in only one cell type, or individual map units significantly enriched for specific GO terms).

The classifications presented here are necessarily limited by the assays and cell lines studied, and probably contain a number of heterogeneous classes of elements. Nonetheless, robust classifications can be made, allowing a systematic view of the human genome.

## Insights into human genomic variation

We next explored the potential impact of sequence variation on ENCODE functional elements. We examined allele-specific variation using results from the GM12878 cells that are derived from an individual (NA12878) sequenced in the 1000 Genomes project, along with her parents. Because ENCODE assays are predominantly sequence-based, the trio design allows each GM12878 data set to be divided by the specific parental contributions at heterozygous sites, producing aggregate haplotypic signals from multiple genomic sites. We examined 193 ENCODE assays for allele-specific biases using 1,409,992 phased, heterozygous SNPs and 167,096 insertions/deletions (indels) (Fig. 8). Alignment biases towards alleles present in the reference genome sequence were avoided using a sequence specifically tailored to the variants and haplotypes present in NA12878 (a 'personalized genome')[72]. We found instances of preferential binding towards each parental allele. For example, comparison of the results from the POLR2A, H3K79me2 and H3K27me3 assays in the region of *NACC2* (Fig. 8a) shows a strong paternal bias for H3K79me2 and POL2RA and a strong maternal bias for H3K27me3, indicating differential activity for the maternal and paternal alleles.

Figure 8b shows the correlation of selected allele-specific signals across the whole genome. For instance, we found a strong allelic correlation between POL2RA and BCLAF1 binding, as well as negative correlation between H3K79me2 and H3K27me3, both at genes (Fig. 8b, below the diagonal, bottom left) and chromosomal segments (top right). Overall, we found that positive allelic correlations among the 193 ENCODE assays are stronger and more frequent than negative correlations. This may be due to preferential capture of accessible alleles and/or the specific histone modification and transcription factor, assays used in the project.

## Rare variants, individual genomes and somatic variants

We further investigated the potential functional effects of individual variation in the context of ENCODE annotations. We divided NA12878 variants into common and rare classes, and partitioned these into those overlapping ENCODE annotation (Fig. 9a and Supplementary Tables 1 and 2, section K). We also predicted potential functional effects: for protein-coding genes, these are either non-synonymous SNPs or variants likely to induce loss of function by frame-shift, premature stop, or splice-site disruption; for other regions, these are variants that overlap a transcription-factor-binding site. We found similar numbers of potentially functional variants affecting protein-coding genes or affecting other ENCODE annotations, indicating that many functional variants within individual genomes lie outside exons of protein-coding genes. A more detailed analysis of regulatory variant annotation is described in ref. 73.

To study further the potential effects of NA12878 genome variants on transcription-factor-binding regions, we performed peak calling using a constructed personal diploid genome sequence for NA12878 (ref. 72). We aligned ChIP-seq sequences from GM12878 separately against the maternal and paternal haplotypes. As expected, a greater fraction of reads were aligned than to the reference genome (see Supplementary Information, Supplementary Fig. 1, section K). On average, approximately 1% of transcription-factor-binding sites in GM12878 cells are detected in a haplotype-specific fashion. For instance, Fig. 9b shows a CTCF-binding site not detected using the

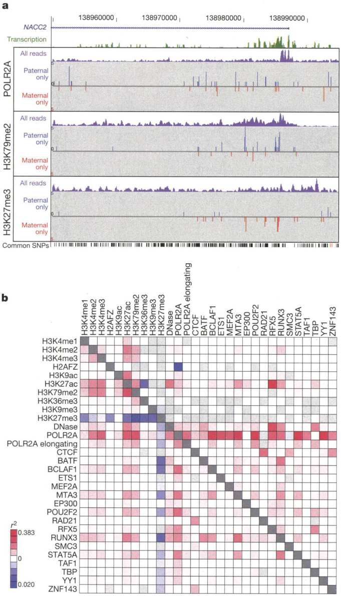


**Figure 8 | Allele-specific ENCODE elements. a,** Representative allele-specific information from GM12878 cells for selected assays around the first exon of the *NACC2* gene (genomic region Chr9: 138950000–138995000, GRCh37). Transcription signal is shown in green, and the three sections show allele-specific data for three data sets (POLR2A, H3K79me2 and H3K27me3 ChIP-seq). In each case the purple signal is the processed signal for all sequence reads for the assay, whereas the blue and red signals show sequence reads specifically assigned to either the paternal or maternal copies of the genome, respectively. The set of common SNPs from dbSNP, including the phased, heterozygous SNPs used to provide the assignment, are shown at the bottom of the panel. *NACC2* has a statistically significant paternal bias for POLR2A and the transcription-associated mark H3K79me2, and has a significant maternal bias for the repressive mark H3K27me3. **b,** Pair-wise correlations of allele-specific signal within single genes (below the diagonal) or within individual ChromHMM segments across the whole genome for selected DNase-seq and histone modification and transcription factor ChIP-seq assays. The extent of correlation is coloured according to the heat-map scale indicated from positive correlation (red) through to anti-correlation (blue). An interactive version of this figure is available in the online version of the paper.

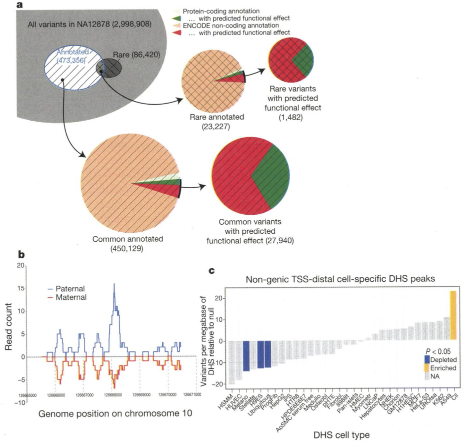


**Figure 9 | Examining ENCODE elements on a per individual basis in the normal and cancer genome.** **a**, Breakdown of variants in a single genome (NA12878) by both frequency (common or rare (that is, variants not present in the low-coverage sequencing of 179 individuals in the pilot 1 European panel of the 1000 Genomes project[55])) and by ENCODE annotation, including protein-coding gene and non-coding elements (GENCODE annotations for protein-coding genes, pseudogenes and other ncRNAs, as well as transcription-factor-binding sites from ChIP-seq data sets, excluding broad annotations such as histone modifications, segmentations and RNA-seq). Annotation status is further subdivided by predicted functional effect, being non-synonymous and missense mutations for protein-coding regions and variants overlapping bound transcription factor motifs for non-coding element annotations. A substantial proportion of variants are annotated as having predicted functional effects in the non-coding category. **b**, One of several relatively rare occurrences, where alignment to an individual genome sequence (paternal and maternal panels) shows a different readout from the reference genome. In this case, a paternal-haplotype-specific CTCF peak is identified. **c**, Relative level of somatic variants from a whole-genome melanoma sample that occur in DHSs unique to different cell lines. The coloured bars show cases that are significantly enriched or suppressed in somatic mutations. Details of ENCODE cell types can be found at http://encodeproject.org/ENCODE/cellTypes.html. An interactive version of this figure is available in the online version of the paper.

reference sequence that is only present on the paternal haplotype due to a 1-bp deletion (see also Supplementary Fig. 2, section K). As costs of DNA sequencing decrease further, optimized analysis of ENCODE-type data should use the genome sequence of the individual or cell being analysed when possible.

Most analyses of cancer genomes so far have focused on characterizing somatic variants in protein-coding regions. We intersected four available whole-genome cancer data sets with ENCODE annotations (Fig. 9c and Supplementary Fig. 2, section L). Overall, somatic variation is relatively depleted from ENCODE annotated regions, particularly for elements specific to a cell type matching the putative tumour source (for example, skin melanocytes for melanoma). Examining the mutational spectrum of elements in introns for cases where a strand-specific mutation assignment could be made reveals that there are mutational spectrum differences between DHSs and unannotated regions (0.06 Fisher's exact test, Supplementary Fig. 3, section L). The suppression of somatic mutation is consistent with important functional roles of these elements within tumour cells, highlighting a potential alternative set of targets for examination in cancer.

## Common variants associated with disease

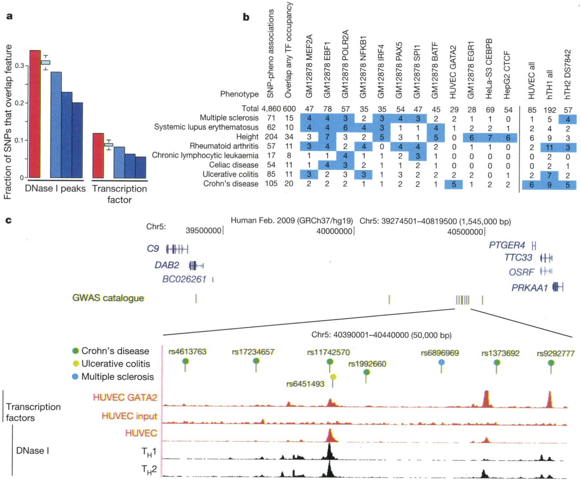
In recent years, GWAS have greatly extended our knowledge of genetic loci associated with human disease risk and other phenotypes. The output of these studies is a series of SNPs (GWAS SNPs) correlated with a phenotype, although not necessarily the functional variants. Notably, 88% of associated SNPs are either intronic or intergenic[74]. We examined 4,860 SNP–phenotype associations for 4,492 SNPs curated in the National Human Genome Research Institute (NHGRI) GWAS catalogue[74]. We found that 12% of these SNPs overlap transcription-factor-occupied regions whereas 34% overlap DHSs (Fig. 10a). Both figures reflect significant enrichments relative to the overall proportions of 1000 Genomes project SNPs (about 6% and 23%, respectively). Even after accounting for biases introduced by selection of SNPs for the standard genotyping arrays, GWAS SNPs show consistently higher overlap with ENCODE annotations (Fig. 10a, see Supplementary Information). Furthermore, after partitioning the genome by density of different classes of functional elements, GWAS SNPs were consistently enriched beyond all the genotyping SNPs in function-rich partitions, and depleted in function-poor partitions (see Supplementary Fig. 1, section M). GWAS SNPs are particularly enriched in the segmentation classes associated with enhancers and TSSs across several cell types (see Supplementary Fig. 2, section M).

Examining the SOM of integrated ENCODE annotations (see above), we found 19 SOM map units showing significant enrichment for GWAS SNPs, including many SOM units previously associated with specific gene functions, such as the immune response regions.

Thus, an appreciable proportion of SNPs identified in initial GWAS scans are either functional or lie within the length of an ENCODE annotation (~500 bp on average) and represent plausible candidates for the functional variant. Expanding the set of feasible functional SNPs to those in reasonable linkage disequilibrium, up to 71% of GWAS SNPs have a potential causative SNP overlapping a DNase I site, and 31% of loci have a candidate SNP that overlaps a binding site occupied by a transcription factor (see also refs 73, 75).

The GWAS catalogue provides a rich functional categorization from the precise phenotypes being studied. These phenotypic categorizations are nonrandomly associated with ENCODE annotations and there is marked correspondence between the phenotype and the identity of the cell type or transcription factor used in the ENCODE assay (Fig. 10b). For example, five SNPs associated with Crohn's disease overlap GATA2-binding sites ($P$ value 0.003 by random permutation or 0.001 by an empirical approach comparing to the GWAS-matched SNPs; see Supplementary Information), and fourteen are located in DHSs found in immunologically relevant cell types. A notable example is a gene desert on chromosome 5p13.1 containing eight SNPs associated with inflammatory diseases. Several are close to or within DHSs in T-helper type 1 ($T_H1$) and $T_H2$ cells as well as peaks of binding by transcription factors in HUVECs (Fig. 10c). The latter cell line is not immunological, but factor occupancy detected there could be a proxy for binding of a more relevant factor, such as GATA3, in T cells. Genetic variants in this region also affect expression levels of *PTGER4* (ref. 76), encoding the prostaglandin receptor EP4. Thus, the ENCODE data reinforce the hypothesis that genetic variants in 5p13.1 modulate the expression of flanking genes, and furthermore provide the specific hypothesis that the variants affect occupancy of a GATA factor in an allele-specific manner, thereby influencing susceptibility to Crohn's disease.

Nonrandom association of phenotypes with ENCODE cell types strengthens the argument that at least some of the GWAS lead SNPs are functional or extremely close to functional variants. Each of the associations between a lead SNP and an ENCODE annotation remains a credible hypothesis of a particular functional element



| Phenotype | SNP-pheno associations | Overlap any TF occupancy | GM12878 MEF2A | GM12878 EBF1 | GM12878 POLR2A | GM12878 NFKB1 | GM12878 IRF4 | GM12878 PAX5 | GM12878 SPI1 | GM12878 BATF | HUVEC GATA2 | GM12878 EGR1 | HeLa-S3 CEBPB | HepG2 CTCF | HUVEC all | hTH1 all | hTH2 DS7842 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Total | 4,860 | 600 | 47 | 78 | 57 | 35 | 35 | 54 | 47 | 45 | 29 | 28 | 69 | 54 | 85 | 192 | 57 |
| Multiple sclerosis | 71 | 15 | 4 | 4 | 3 | 2 | 3 | 4 | 3 | 2 | 2 | 1 | 1 | 0 | 1 | 5 | 4 |
| Systemic lupus erythematosus | 62 | 10 | 4 | 4 | 6 | 4 | 3 | 1 | 1 | 4 | 1 | 2 | 2 | 1 | 2 | 4 | 2 |
| Height | 204 | 34 | 3 | 7 | 3 | 2 | 5 | 3 | 1 | 5 | 0 | 6 | 7 | 6 | 6 | 9 | 3 |
| Rheumatoid arthritis | 57 | 11 | 3 | 4 | 2 | 4 | 0 | 4 | 4 | 2 | 0 | 1 | 1 | 0 | 2 | 11 | 3 |
| Chronic lymphocytic leukaemia | 17 | 8 | 1 | 1 | 4 | 1 | 1 | 2 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 0 |
| Celiac disease | 54 | 11 | 1 | 4 | 3 | 2 | 1 | 1 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 1 |
| Ulcerative colitis | 85 | 11 | 3 | 2 | 3 | 3 | 1 | 3 | 2 | 2 | 2 | 2 | 0 | 1 | 2 | 7 | 2 |
| Crohn's disease | 105 | 20 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 5 | 1 | 2 | 2 | 6 | 9 | 5 |

**Figure 10 | Comparison of genome-wide-association-study-identified loci with ENCODE data.** **a**, Overlap of lead SNPs in the NHGRI GWAS SNP catalogue (June 2011) with DHSs (left) or transcription-factor-binding sites (right) as red bars compared with various control SNP sets in blue. The control SNP sets are (from left to right): SNPs on the Illumina 2.5M chip as an example of a widely used GWAS SNP typing panel; SNPs from the 1000 Genomes project; SNPs extracted from 24 personal genomes (see personal genome variants track at http://main.genome-browser.bx.psu.edu (ref. 80)), all shown as blue bars. In addition, a further control used 1,000 randomizations from the genotyping SNP panel, matching the SNPs with each NHGRI catalogue SNP for allele frequency and distance to the nearest TSS (light blue bars with bounds at 1.5 times the interquartile range). For both DHSs and transcription-factor-binding regions, a larger proportion of overlaps with GWAS-implicated SNPs is found compared to any of the controls sets. **b**, Aggregate overlap of phenotypes to selected transcription-factor-binding sites (left matrix) or DHSs in selected cell lines (right matrix), with a count of overlaps between the phenotype and the cell line/factor. Values in blue squares pass an empirical $P$-value threshold $\leq 0.01$ (based on the same analysis of overlaps between randomly chosen, GWAS-matched SNPs and these epigenetic features) and have at least a count of three overlaps. The $P$ value for the total number of phenotype–transcription factor associations is $<0.001$. **c**, Several SNPs associated with Crohn's disease and other inflammatory diseases that reside in a large gene desert on chromosome 5, along with some epigenetic features indicative of function. The SNP (rs11742570) strongly associated to Crohn's disease overlaps a GATA2 transcription-factor-binding signal determined in HUVECs. This region is also DNase I hypersensitive in HUVECs and T-helper $T_H1$ and $T_H2$ cells. An interactive version of this figure is available in the online version of the paper.

class or cell type to explore with future experiments. Supplementary Tables 1–3, section M, list all 14,885 pairwise associations across the ENCODE annotations. The accompanying papers have a more detailed examination of common variants with other regulatory information[19,25,29,73,75,77].

## Concluding remarks

The unprecedented number of functional elements identified in this study provides a valuable resource to the scientific community as well as significantly enhances our understanding of the human genome. Our analyses have revealed many novel aspects of gene expression and regulation as well as the organization of such information, as illustrated by the accompanying papers (see http://www.encodeproject.org/ENCODE/pubs.html for collected ENCODE publications). However, there are still many specific details, particularly about the mechanistic processes that generate these elements and how and where they function, that require additional experiments to elucidate.

The large spread of coverage—from our highest resolution, most conservative set of bases implicated in GENCODE protein-coding gene exons (2.9%) or specific protein DNA binding (8.5%) to the broadest, most general set of marks covering the genome (approximately 80%), with many gradations in between—presents a spectrum of elements with different functional properties discovered by ENCODE. A total of 99% of the known bases in the genome are within 1.7 kb of any ENCODE element, whereas 95% of bases are within 8 kb of a bound transcription factor motif or DNase I footprint. Interestingly, even using the most conservative estimates, the fraction of bases likely to be involved in direct gene regulation, even though incomplete, is significantly higher than that ascribed to protein-coding exons (1.2%), raising the possibility that more information in the human genome may be important for gene regulation than for biochemical function. Many of the regulatory elements are not constrained across mammalian evolution, which so far has been one of the most reliable indications of an important biochemical event for the organism. Thus, our data provide orthologous indicators for suggesting possible functional elements.

Importantly, for the first time we have sufficient statistical power to assess the impact of negative selection on primate-specific elements, and all ENCODE classes display evidence of negative selection in these unique-to-primate elements. Furthermore, even with our most conservative estimate of functional elements (8.5% of putative DNA/protein binding regions) and assuming that we have already sampled half of the elements from our transcription factor and cell-type diversity, one would estimate that at a minimum 20% (17% from protein binding and 2.9% protein coding gene exons) of the genome participates in these specific functions, with the likely figure significantly higher.

The broad coverage of ENCODE annotations enhances our understanding of common diseases with a genetic component, rare genetic diseases, and cancer, as shown by our ability to link otherwise anonymous associations to a functional element. ENCODE and similar studies provide a first step towards interpreting the rest of the genome—beyond protein-coding genes—thereby augmenting common disease genetic studies with testable hypotheses. Such information justifies performing whole-genome sequencing (rather than exome only, 1.2% of the genome) on rare diseases and investigating somatic variants in non-coding functional elements, for instance, in cancer. Furthermore, as GWAS analyses typically associate disease to SNPs in large regions, comparison to ENCODE non-coding functional elements can help pinpoint putative causal variants in addition to refinement of location by fine-mapping techniques[78]. Combining ENCODE data with allele-specific information derived from individual genome sequences provides specific insight on the impact of a genetic variant. Indeed, we believe that a significant goal would be to use functional data such as that derived from this project to assign every genomic variant to its possible impact on human phenotypes.

So far, ENCODE has sampled 119 of 1,800 known transcription factors and general components of the transcriptional machinery on a limited number of cell types, and 13 of more than 60 currently known histone or DNA modifications across 147 cell types. DNase I, FAIRE and extensive RNA assays across subcellular fractionations have been undertaken on many cell types, but overall these data reflect a minor fraction of the potential functional information encoded in the human genome. An important future goal will be to enlarge this data set to additional factors, modifications and cell types, complementing the other related projects in this area (for example, Roadmap Epigenomics Project, http://www.roadmapepigenomics.org/, and International Human Epigenome Consortium, http://www.ihec-epigenomes.org/). These projects will constitute foundational resources for human genomics, allowing a deeper interpretation of the organization of gene and regulatory information and the mechanisms of regulation, and thereby provide important insights into human health and disease. Co-published ENCODE-related papers can be explored online via the *Nature* ENCODE explorer (http://www.nature.com/ENCODE), a specially designed visualization tool that allows users to access the linked papers and investigate topics that are discussed in multiple papers via thematically organized threads.

## METHODS SUMMARY

For full details of Methods, see Supplementary Information.

**Received 24 November 2011; accepted 29 May 2012.**

1. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306,** 636–640 (2004).
2. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447,** 799–816 (2007).
3. The ENCODE Project Consortium. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* **9,** e1001046 (2011).
4. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420,** 520–562 (2002).
5. Chiaromonte, F. *et al.* The share of human genomic DNA under selection estimated from human-mouse genomic alignments. *Cold Spring Harb. Symp. Quant. Biol.* **68,** 245–254 (2003).
6. Cooper, G. M. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* **15,** 901–913 (2005).
7. Parker, S. C., Hansen, L., Abaan, H. O., Tullius, T. D. & Margulies, E. H. Local DNA topography correlates with functional noncoding regions of the human genome. *Science* **324,** 389–392 (2009).
8. Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478,** 476–482 (2011).
9. Pheasant, M. & Mattick, J. S. Raising the estimate of functional human sequences. *Genome Res.* **17,** 1245–1253 (2007).
10. Ponting, C. P. & Hardison, R. C. What fraction of the human genome is functional? *Genome Res.* **21,** 1769–1776 (2011).
11. Asthana, S. *et al.* Widely distributed noncoding purifying selection in the human genome. *Proc. Natl Acad. Sci. USA* **104,** 12410–12415 (2007).
12. Landt, S. G. *et al.* ChIP-seq guidelines and practices used by the ENCODE and modENCODE consortia. *Genome Res.* http://dx.doi.org/10.1101/gr.136184.111 (2012).
13. Li, Q., Brown, J. B., Huang, H. & Bickel, P. J. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5,** 1752–1779 (2011).
14. Harrow, J. *et al.* GENCODE: The reference human genome annotation for the ENCODE project. *Genome Res.* http://dx.doi.org/10.1101/gr.135350.111 (2012).
15. Howald, C. *et al.* Combining RT-PCR-seq and RNA-seq to catalog all genic elements encoded in the human genome. *Genome Res.* http://dx.doi.org/10.1101/gr.134478.111 (2012).
16. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* http://dx.doi.org/10.1038/nature11233 (this issue).
17. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* http://dx.doi.org/10.1101/gr.132159.111 (2012).
18. Pei, B. *et al.* The GENCODE pseudogene resource. *Genome Biol.* **13,** R51 (2012).
19. Gerstein, M. B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* http://dx.doi.org/10.1038/nature11245 (this issue).
20. Bickel, P. J., Boley, N., Brown, J. B., Huang, H. Y. & Zhang, N. R. Subsampling methods for genomic inference. *Ann. Appl. Stat.* **4,** 1660–1697 (2010).
21. Kaplan, T. *et al.* Quantitative models of the mechanisms that control genome-wide patterns of transcription factor binding during early *Drosophila* development. *PLoS Genet.* **7,** e1001290 (2011).
22. Li, X. Y. *et al.* The role of chromatin accessibility in directing the widespread, overlapping patterns of *Drosophila* transcription factor binding. *Genome Biol.* **12,** R34 (2011).

23. Pique-Regi, R. *et al.* Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res.* **21,** 447–455 (2011).
24. Zhang, Y. *et al.* Primary sequence and epigenetic determinants of *in vivo* occupancy of genomic DNA by GATA1. *Nucleic Acids Res.* **37,** 7024–7038 (2009).
25. Neph, S. *et al.* An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* http://dx.doi.org/10.1038/nature11212 (this issue).
26. Whitfield, T. W. *et al.* Functional analysis of transcription factor binding sites in human promoters. *Genome Biol.* **13,** R50 (2012).
27. Gross, D. S. & Garrard, W. T. Nuclease hypersensitive sites in chromatin. *Annu. Rev. Biochem.* **57,** 159–197 (1988).
28. Urnov, F. D. Chromatin remodeling as a guide to transcriptional regulatory networks in mammals. *J. Cell. Biochem.* **88,** 684–694 (2003).
29. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* http://dx.doi.org/10.1038/nature11232 (this issue).
30. Kundaje, A. *et al.* Ubiquitous heterogeneity and asymmetry of the chromatin environment at regulatory elements. *Genome Res.* http://dx.doi.org/10.1101/gr.136366.111 (2012).
31. Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G. & Rauscher, F. J. III. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev.* **16,** 919–932 (2002).
32. Frietze, S., O'Geen, H., Blahnik, K. R., Jin, V. X. & Farnham, P. J. ZNF274 recruits the histone methyltransferase SETDB1 to the 3′ ends of ZNF genes. *PLoS ONE* **5,** e15082 (2010).
33. Boyle, A. P. *et al.* High-resolution genome-wide *in vivo* footprinting of diverse transcription factors in human cells. *Genome Res.* **21,** 456–464 (2011).
34. Hesselberth, J. R. *et al.* Global mapping of protein-DNA interactions *in vivo* by digital genomic footprinting. *Nature Methods* **6,** 283–289 (2009).
35. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).
36. Kouzarides, T. Chromatin modifications and their function. *Cell* **128,** 693–705 (2007).
37. Li, B., Carey, M. & Workman, J. L. The role of chromatin during transcription. *Cell* **128,** 707–719 (2007).
38. Hon, G. C., Hawkins, R. D. & Ren, B. Predictive chromatin signatures in the mammalian genome. *Hum. Mol. Genet.* **18,** R195–R201 (2009).
39. Zhou, V. W., Goren, A. & Bernstein, B. E. Charting histone modifications and the functional organization of mammalian genomes. *Nature Rev. Genet.* **12,** 7–18 (2011).
40. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473,** 43–49 (2011).
41. Hon, G., Wang, W. & Ren, B. Discovery and annotation of functional chromatin signatures in the human genome. *PLoS Comput. Biol.* **5,** e1000566 (2009).
42. Ball, M. P. *et al.* Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nature Biotechnol.* **27,** 361–368 (2009).
43. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454,** 766–770 (2008).
44. Ogryzko, V. V., Schiltz, R. L., Russanova, V., Howard, B. H. & Nakatani, Y. The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* **87,** 953–959 (1996).
45. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462,** 315–322 (2009).
46. Dekker, J. Gene regulation in the third dimension. *Science* **319,** 1793–1794 (2008).
47. Dostie, J. *et al.* Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* **16,** 1299–1309 (2006).
48. Lajoie, B. R., van Berkum, N. L., Sanyal, A. & Dekker, J. My5C: web tools for chromosome conformation capture studies. *Nature Methods* **6,** 690–691 (2009).
49. Sanyal, A., Lajoie, B., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* http://dx.doi.org/10.1038/nature11279 (this issue).
50. Fullwood, M. J. *et al.* An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462,** 58–64 (2009).
51. Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148,** 84–98 (2012).
52. Borneman, A. R. *et al.* Divergence of transcription factor binding sites across related yeast species. *Science* **317,** 815–819 (2007).
53. Odom, D. T. *et al.* Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nature Genet.* **39,** 730–732 (2007).
54. Schmidt, D. *et al.* Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328,** 1036–1040 (2010).
55. A map of human genome variation from population-scale sequencing. *Nature* **467,** 1061–1073 (2010).
56. King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188,** 107–116 (1975).
57. Spivakov, M. *et al.* Analysis of variation at transcription factor binding sites in *Drosophila* and humans. *Genome Biol.* **13,** R49 (2012).
58. Sandelin, A. *et al.* Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nature Rev. Genet.* **8,** 424–436 (2007).
59. Dong, X. *et al.* Modeling gene expression using chromatin features in various cellular contexts. *Genome Biol.* **13,** R53 (2012).
60. Huff, J. T., Plocik, A. M., Guthrie, C. & Yamamoto, K. R. Reciprocal intronic and exonic histone modification regions in humans. *Nature Struct. Mol. Biol.* **17,** 1495–1499 (2010).
61. Tilgner, H. *et al.* Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* http://dx.doi.org/10.1101/gr.134445.111 (2012).
62. Fu, Y., Sinha, M., Peterson, C. L. & Weng, Z. The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet.* **4,** e1000138 (2008).
63. Kornberg, R. D. & Stryer, L. Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. *Nucleic Acids Res.* **16,** 6677–6690 (1988).
64. Schones, D. E. *et al.* Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132,** 887–898 (2008).
65. Valouev, A. *et al.* Determinants of nucleosome organization in primary human cells. *Nature* **474,** 516–520 (2011).
66. Frietze, S. *et al.* Cell type-specific binding patterns reveal that TCF7L2 can be tethered to the genome by association with GATA3. *Genome Biol.* **13,** R52 (2012).
67. Yip, K. Y. *et al.* Classification of human genomic regions based on experimentally-determined binding sites of more than 100 transcription-related factors. *Genome Biol.* **13,** R48 (2012).
68. Hoffman, M. M. *et al.* Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nature Methods* **9,** 473–476 (2012).
69. Kapranov, P. *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316,** 1484–1488 (2007).
70. Koch, F. *et al.* Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nature Struct. Mol. Biol.* **18,** 956–963 (2011).
71. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnol.* **28,** 495–501 (2010).
72. Rozowsky, J. *et al.* AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol. Syst. Biol.* **7,** 522 (2011).
73. Boyle, A. P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* http://dx.doi.org/10.1101/gr.137323.112 (2012).
74. Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106,** 9362–9367 (2009).
75. Schaub, M. A., Boyle, A. P., Kundaje, A., Batzoglou, S. & Snyder, M. Linking disease associations with regulatory information in the human genome. *Genome Res.* http://dx.doi.org/10.1101/gr.136127.111 (2012).
76. Libioulle, C. *et al.* Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLoS Genet.* **3,** e58 (2007).
77. Vernot, B. *et al.* Personal and population genomics of human regulatory variation. *Genome Res.* http://dx.doi.org/10.1101/gr.134890.111 (2012).
78. Harismendy, O. *et al.* 9p21 DNA variants associated with coronary artery disease impair interferon-$\gamma$ signalling response. *Nature* **470,** 264–268 (2011).
79. Cheng, C. *et al.* Understanding transcriptional regulation by integrative analysis of transcription factor binding data. *Genome Res.* http://dx.doi.org/10.1101/gr.136838.111 (2012).
80. Schuster, S. C. *et al.* Complete Khoisan and Bantu genomes from southern Africa. *Nature* **463,** 943–947 (2010).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank additional members of our laboratories and institutions who have contributed to the experimental and analytical components of this project. We thank D. Leja for assistance with production of the figures. The Consortium is funded by grants from the NHGRI as follows: production grants: U54HG004570 (B. E. Bernstein); U01HG004695 (E. Birney); U54HG004563 (G. E. Crawford); U54HG004557 (T. R. Gingeras); U54HG004555 (T. J. Hubbard); U41HG004568 (W. J. Kent); U54HG004576 (R. M. Myers); U54HG004558 (M. Snyder); U54HG004592 (J. A. Stamatoyannopoulos). Pilot grants: R01HG003143 (J. Dekker); RC2HG005591 and R01HG003700 (M. C. Giddings); R01HG004456-03 (Y. Ruan); U01HG004571 (S. A. Tenenbaum); U01HG004561 (Z. Weng); RC2HG005679 (K. P. White). This project was supported in part by American Recovery and Reinvestment Act (ARRA) funds from the NHGRI through grants U54HG004570, U54HG004563, U41HG004568, U54HG004592, R01HG003143, RC2HG005591, R01HG003541, U01HG004561, RC2HG005679 and R01HG003988 (L. Pennacchio). In addition, work from NHGRI Groups was supported by the Intramural Research Program of the NHGRI (L. Elnitski, ZIAHG200323; E. H. Margulies, ZIAHG200341). Research in the Pennachio laboratory was performed at Lawrence Berkeley National Laboratory and at the United States Department of Energy Joint Genome Institute, Department of Energy Contract DE-AC02-05CH11231, University of California.

**Author Contributions** See the consortium author list for details of author contributions.

**Author Information** The Supplementary Information is accompanied by a Virtual Machine (VM) containing the functioning analysis data and code. Further details of the VM are available from http://encodeproject.org/ENCODE/integrativeAnalysis/VM. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and the online version of the paper is freely available to all readers. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.B. (birney@ebi.ac.uk).

## The ENCODE Project Consortium

**Overall coordination (data analysis coordination)** Ian Dunham[1], Anshul Kundaje[2]†; **Data production leads (data production)** Shelley F. Aldred[3], Patrick J. Collins[3], Carrie A. Davis[4], Francis Doyle[5], Charles B. Epstein[6], Seth Frietze[7], Jennifer Harrow[8], Rajinder Kaul[9], Jainab Khatun[10], Bryan R. Lajoie[11], Stephen G. Landt[12], Bum-Kyu Lee[13], Florencia Pauli[14], Kate R. Rosenbloom[15], Peter Sabo[16], Alexias Safi[17], Amartya Sanyal[11], Noam Shoresh[6], Jeremy M. Simon[18], Lingyun Song[17], Nathan D. Trinklein[3]; **Lead analysts (data analysis)** Robert C. Altshuler[19], Ewan Birney[1], James B. Brown[20], Chao Cheng[21], Sarah Djebali[22], Xianjun Dong[23], Ian Dunham[1], Jason Ernst[19]†, Terrence S. Furey[24], Mark Gerstein[21], Belinda Giardine[25], Melissa Greven[23], Ross C. Hardison[25,26], Robert S. Harris[25], Javier Herrero[1], Michael M. Hoffman[16], Sowmya Iyer[27], Manolis Kellis[19], Jainab Khatun[10], Pouya Kheradpour[19], Anshul Kundaje[2]†, Timo Lassmann[28], Qunhua Li[20]†, Xinying Lin[23], Georgi K. Marinov[29], Angelika Merkel[22], Ali Mortazavi[30], Stephen C. J. Parker[31], Timothy E. Reddy[14]†, Joel Rozowsky[21], Felix Schlesinger[4], Robert E. Thurman[16], Jie Wang[23], Lucas D. Ward[19], Troy W. Whitfield[23], Steven P. Wilder[1], Weisheng Wu[25], Hualin S. Xi[32], Kevin Y. Yip[21]†, Jiali Zhuang[23]; **Writing group** Bradley E. Bernstein[6,33], Ewan Birney[1], Ian Dunham[1], Eric D. Green[34], Chris Gunter[14], Michael Snyder[12]; **NHGRI project management (scientific management)** Michael J. Pazin[35], Rebecca F. Lowdon[35]†, Laura A. L. Dillon[35]†, Leslie B. Adams[35], Caroline J. Kelly[35], Julia Zhang[35]†, Judith R. Wexler[35]†, Eric D. Green[34], Peter J. Good[35], Elise A. Feingold[35]; **Principal investigators (steering committee)** Bradley E. Bernstein[6,33], Ewan Birney[1], Gregory E. Crawford[17,36], Job Dekker[11], Laura Elnitski[37], Peggy J. Farnham[7], Mark Gerstein[21], Morgan C. Giddings[10], Thomas R. Gingeras[4,38], Eric D. Green[34], Roderic Guigó[22,39], Ross C. Hardison[25,26], Timothy J. Hubbard[8], Manolis Kellis[19], W. James Kent[15], Jason D. Lieb[18], Elliott H. Margulies[31]†, Richard M. Myers[14], Michael Snyder[12], John A. Stamatoyannopoulos[40], Scott A. Tenenbaum[5], Zhiping Weng[23], Kevin P. White[41], Barbara Wold[29,42]; **Boise State University and University of North Carolina at Chapel Hill Proteomics groups (data production and analysis)** Jainab Khatun[10], Yanbao Yu[43], John Wrobel[10], Brian A. Risk[10], Harsha P. Gunawardena[43], Heather C. Kuiper[43], Christopher W. Maier[43], Ling Xie[43], Xian Chen[43], Morgan C. Giddings[10]; **Broad Institute Group (data production and analysis)** Bradley E. Bernstein[6,33], Charles B. Epstein[6], Noam Shoresh[6], Jason Ernst[19]†, Pouya Kheradpour[19], Tarjei S. Mikkelsen[6], Shawn Gillespie[33], Alon Goren[6,33], Oren Ram[6,33], Xiaolan Zhang[6], Li Wang[6], Robbyn Issner[6], Michael J. Coyne[6], Timothy Durham[6], Manching Ku[6,33], Thanh Truong[6], Lucas D. Ward[19], Robert C. Altshuler[19], Matthew L. Eaton[19], Manolis Kellis[19]; **Cold Spring Harbor, University of Geneva, Center for Genomic Regulation, Barcelona, RIKEN, Sanger Institute, University of Lausanne, Genome Institute of Singapore group (data production and analysis)** Sarah Djebali[22], Carrie A. Davis[4], Angelika Merkel[22], Alex Dobin[4], Timo Lassmann[28], Ali Mortazavi[30], Andrea Tanzer[22], Julien Lagarde[22], Wei Lin[4], Felix Schlesinger[4], Chenghai Xue[4], Georgi K. Marinov[29], Jainab Khatun[10], Brian A. Williams[29], Chris Zaleski[4], Joel Rozowsky[21], Maik Röder[22], Felix Kokocinski[8]†, Rehab F. Abdelhamid[28], Tyler Alioto[22,44], Igor Antoshechkin[29], Michael T. Baer[4], Philippe Batut[4], Ian Bell[45], Kimberly Bell[4], Sudipto Chakrabortty[4], Xian Chen[43], Jacqueline Chrast[46], Joao Curado[22], Thomas Derrien[22]†, Jorg Drenkow[4], Erica Dumais[45], Jackie Dumais[45], Radha Duttagupta[45], Megan Fastuca[4], Kata Fejes-Toth[4]†, Pedro Ferreira[22], Sylvain Foissac[45], Melissa J. Fullwood[47]†, Hui Gao[45], David Gonzalez[22], Assaf Gordon[4], Harsha P. Gunawardena[43], Cédric Howald[46], Sonali Jha[4], Rory Johnson[22], Philipp Kapranov[45]†, Brandon King[29], Colin Kingswood[22,44], Guoliang Li[48], Oscar J. Luo[47], Eddie Park[30], Jonathan B. Preall[4], Kimberly Presaud[4], Paolo Ribeca[22,44], Brian A. Risk[10], Daniel Robyr[49], Xiaoan Ruan[47], Michael Sammeth[22,44], Kuljeet Singh Sandhu[47], Lorain Schaeffer[29], Lei-Hoon See[4], Atif Shahab[47], Jorgen Skancke[22], Ana Maria Suzuki[28], Hazuki Takahashi[28], Hagen Tilgner[22]†, Diane Trout[29], Nathalie Walters[46], Huaien Wang[4], John Wrobel[10], Yanbao Yu[43], Yoshihide Hayashizaki[28], Jennifer Harrow[8], Mark Gerstein[21], Timothy J. Hubbard[8], Alexandre Reymond[46], Stylianos E. Antonarakis[49], Gregory J. Hannon[4], Morgan C. Giddings[10], Yijun Ruan[47], Barbara Wold[29,42], Piero Carninci[28], Roderic Guigó[22,39], Thomas R. Gingeras[4,38]; **Data coordination center at UC Santa Cruz (production data coordination)** Kate R. Rosenbloom[15], Cricket A. Sloan[15], Katrina Learned[15], Venkat S. Malladi[15], Matthew C. Wong[15], Galt P. Barber[15], Melissa S. Cline[15], Timothy R. Dreszer[15], Steven G. Heitner[15], Donna Karolchik[15], W. James Kent[15], Vanessa M. Kirkup[15], Laurence R. Meyer[15], Jeffrey C. Long[15], Morgan Maddren[15], Brian J. Raney[15]; **Duke University, EBI, University of Texas, Austin, University of North Carolina-Chapel Hill group (data production and analysis)** Terrence S. Furey[24], Lingyun Song[17], Linda L. Grasfeder[18], Paul G. Giresi[18], Bum-Kyu Lee[13], Anna Battenhouse[13], Nathan C. Sheffield[17], Jeremy M. Simon[18], Kimberly A. Showers[18], Alexias Safi[17], Darin London[17], Akshay A. Bhinge[13], Christopher Shestak[18], Matthew R. Schaner[18], Seul Ki Kim[18], Zhuzhu Z. Zhang[18], Piotr A. Mieczkowski[50], Joanna O. Mieczkowska[18], Zheng Liu[13], Ryan M. McDaniell[13], Yunyun Ni[13], Naim U. Rashid[51], Min Jae Kim[18], Sheera Adar[18], Zhancheng Zhang[24], Tianyuan Wang[17], Deborah Winter[17], Damian Keefe[1], Ewan Birney[1], Vishwanath R. Iyer[13], Jason D. Lieb[18], Gregory E. Crawford[17,36]; **Genome Institute of Singapore group (data production and analysis)** Guoliang Li[48], Kuljeet Singh Sandhu[47], Meizhen Zheng[47], Ping Wang[47], Oscar J. Luo[47], Atif Shahab[47], Melissa J. Fullwood[47]†, Xiaoan Ruan[47], Yijun Ruan[47]; **HudsonAlpha Institute, Caltech, UC Irvine, Stanford group (data production and analysis)** Richard M. Myers[14], Florencia Pauli[14], Brian A. Williams[29], Jason Gertz[14], Georgi K. Marinov[29], Timothy E. Reddy[14]†, Jost Vielmetter[29,42], E. Christopher Partridge[14], Diane Trout[29], Katherine E. Varley[14], Clarke Gasper[29,42], Anita Bansal[14], Shirley Pepke[29,52], Preti Jain[14], Henry Amrhein[29], Kevin M. Bowling[14], Michael Anaya[29,42], Marie K. Cross[14], Brandon King[29], Michael A. Muratet[14], Igor Antoshechkin[29], Kimberly M. Newberry[14], Kenneth McCue[29], Amy S. Nesmith[14], Katherine I. Fisher-Aylor[29,42], Barbara Pusey[14], Gilberto DeSalvo[29,42], Stephanie L. Parker[14]†, Sreeram Balasubramanian[29,42], Nicholas S. Davis[14], Sarah K. Meadows[14], Tracy Eggleston[14], Chris Gunter[14], J. Scott Newberry[14], Shawn E. Levy[14], Devin M. Absher[14], Ali Mortazavi[30], Wing H. Wong[53], Barbara Wold[29,42]; **Lawrence Berkeley National Laboratory group (targeted experimental validation)** Matthew J. Blow[54], Axel Visel[54,55], Len A. Pennachio[54,55]; **NHGRI groups (data production and analysis)** Laura Elnitski[37], Elliott H. Margulies[31]†, Stephen C. J. Parker[31], Hanna M. Petrykowska[37]; **Sanger Institute, Washington University, Yale University, Center for Genomic Regulation, Barcelona, UCSC, MIT, University of Lausanne, CNIO group (data production and analysis)** Alexej Abyzov[21], Bronwen Aken[8], Daniel Barrell[8], Gemma Barson[8], Andrew Berry[8], Alexandra Bignell[8], Veronika Boychenko[8], Giovanni Bussotti[22], Jacqueline Chrast[46], Claire Davidson[8], Thomas Derrien[22]†, Gloria Despacio-Reyes[8], Mark Diekhans[15], Iakes Ezkurdia[56], Adam Frankish[8], James Gilbert[8], Jose Manuel Gonzalez[8], Ed Griffiths[8], Rachel Harte[15], David A. Hendrix[19], Cédric Howald[46], Toby Hunt[8], Irwin Jungreis[19], Mike Kay[8], Ekta Khurana[21], Felix Kokocinski[8]†, Jing Leng[21], Michael F. Lin[19], Jane Loveland[8], Zhi Lu[57], Deepa Manthravadi[8], Marco Mariotti[22], Jonathan Mudge[8], Gaurab Mukherjee[8], Cedric Notredame[22], Baikang Pei[21], Jose Manuel Rodriguez[56], Gary Saunders[8], Andrea Sboner[58], Stephen Searle[8], Cristina Sisu[21], Catherine Snow[8], Charlie Steward[8], Andrea Tanzer[22], Electra Tapanari[8], Michael L. Tress[56], Marijke J. van Baren[59]†, Nathalie Walters[46], Stefan Washietl[19], Laurens Wilming[8], Amonida Zadissa[8], Zhengdong Zhang[60], Michael Brent[59], David Haussler[61], Manolis Kellis[19], Alfonso Valencia[56], Mark Gerstein[21], Alexandre Reymond[46], Roderic Guigó[22,39], Jennifer Harrow[8], Timothy J. Hubbard[8]; **Stanford-Yale, Harvard, University of Massachusetts Medical School, University of Southern California/UC Davis group (data production and analysis)** Stephen G. Landt[12], Seth Frietze[7], Alexej Abyzov[21], Nick Addleman[12], Roger P. Alexander[21], Raymond K. Auerbach[21], Suganthi Balasubramanian[21], Keith Bettinger[12], Nitin Bhardwaj[21], Alan P. Boyle[12], Alina R. Cao[62], Philip Cayting[12], Alexandra Charos[63], Yong Cheng[12], Chao Cheng[21], Catharine Eastman[12], Ghia Euskirchen[12], Joseph D. Fleming[64], Fabian Grubert[12], Lukas Habegger[21], Manoj Hariharan[12], Arif Harmanci[21], Sushma Iyengar[65], Victor X. Jin[66], Konrad J. Karczewski[12], Maya Kasowski[12], Phil Lacroute[12], Hugo Lam[12], Nathan Lamarre-Vincent[64], Jing Leng[21], Jin Lian[67], Marianne Lindahl-Allen[64], Renqiang Min[21]†, Benoit Miotto[64], Hannah Monahan[63], Zarmik Moqtaderi[64], Xinmeng J. Mu[21], Henriette O'Geen[62], Zhengqing Ouyang[12], Dorrelyn Patacsil[12], Baikang Pei[21], Debasish Raha[63], Lucia Ramirez[12], Brian Reed[63], Joel Rozowsky[21], Andrea Sboner[58], Minyi Shi[12], Cristina Sisu[21], Teri Slifer[12], Heather Witt[7], Linfeng Wu[12], Xiaoqin Xu[62], Koon-Kiu Yan[21], Xinqiong Yang[12], Kevin Y. Yip[21]†, Zhengdong Zhang[60], Kevin Struhl[64], Sherman M. Weissman[67], Mark Gerstein[21], Peggy J. Farnham[7], Michael Snyder[12]; **University of Albany SUNY group (data production and analysis)** Scott A. Tenenbaum[5], Luiz O. Penalva[68], Francis Doyle[5]; **University of Chicago, Stanford group (data production and analysis)** Subhradip Karmakar[41], Stephen G. Landt[12], Raj R. Bhanvadia[41], Alina Choudhury[41], Marc Domanus[41], Lijia Ma[41], Jennifer Moran[41], Dorrelyn Patacsil[12], Teri Slifer[12], Alec Victorsen[41], Xinqiong Yang[12], Michael Snyder[12], Kevin P. White[41]; **University of Heidelberg group (targeted experimental validation)** Thomas Auer[69]†, Lazaro Centanin[69], Michael Eichenlaub[69], Franziska Gruhl[69], Stephan Heermann[69], Burkhard Hoeckendorf[69], Daigo Inoue[69], Tanja Kellner[69], Stephan Kirchmaier[69], Claudia Mueller[69], Robert Reinhardt[69], Lea Schertel[69], Stephanie Schneider[69], Rebecca Sinn[69], Beate Wittbrodt[69], Jochen Wittbrodt[69]; **University of Massachusetts Medical School Bioinformatics group (data production and analysis)** Zhiping Weng[23], Troy W. Whitfield[23], Jie Wang[23], Patrick J. Collins[3], Shelley F. Aldred[3], Nathan D. Trinklein[3], E. Christopher Partridge[14], Richard M. Myers[14]; **University of Massachusetts Medical School Genome Folding group (data production and analysis)** Job Dekker[11], Gaurav Jain[11], Bryan R. Lajoie[11], Amartya Sanyal[11]; **University of Washington, University of Massachusetts Medical Center group (data production and analysis)** Gayathri Balasundaram[70], Daniel L. Bates[16], Rachel Byron[70], Theresa K. Canfield[16], Morgan J. Diegel[16], Douglas Dunn[16], Abigail K. Ebersol[71], Tristan Frum[71], Kavita Garg[72], Erica Gist[16], R. Scott Hansen[71], Lisa Boatman[71], Eric Haugen[16], Richard Humbert[16], Gaurav Jain[11], Audra K. Johnson[16], Ericka M. Johnson[71], Tattyana V. Kutyavin[16], Bryan R. Lajoie[11], Kristen Lee[16], Dimitra Lotakis[71], Matthew T. Maurano[16], Shane J. Neph[16], Fiedencio V. Neri[16], Eric D. Nguyen[71], Hongzhu Qu[16], Alex P. Reynolds[16], Vaughn Roach[16], Eric Rynes[16], Peter Sabo[16], Minerva E. Sanchez[71], Richard S. Sandstrom[16], Amartya Sanyal[11], Anthony O. Shafer[16], Andrew B. Stergachis[16], Sean Thomas[16], Robert E. Thurman[16], Benjamin Vernot[16], Jeff Vierstra[16], Shinny Vong[16], Hao Wang[16], Molly A. Weaver[16], Yongqi Yan[71], Miaohua Zhang[70], Joshua M. Akey[16], Michael Bender[70], Michael O. Dorschner[73], Mark Groudine[70], Michael J. MacCoss[16], Patrick Navas[71], George Stamatoyannopoulos[71], Rajinder Kaul[9], Job Dekker[11], John A. Stamatoyannopoulos[40]; **Data Analysis Center (data analysis)**; Ian Dunham[1], Kathryn Beal[1], Alvis Brazma[74], Paul Flicek[1], Javier Herrero[1], Nathan Johnson[1], Damian Keefe[1], Margus Lukk[74]†, Nicholas M. Luscombe[75], Daniel Sobral[1]†, Juan M. Vaquerizas[75], Steven P. Wilder[1], Serafim Batzoglou[2], Arend Sidow[76], Nadine Hussami[2], Sofia Kyriazopoulou-Panagiotopoulou[2], Max W. Libbrecht[2]†, Marc A. Schaub[2], Anshul Kundaje[2]†, Ross C. Hardison[25,26], Webb Miller[25], Belinda Giardine[25], Robert S. Harris[25], Weisheng Wu[25], Peter J. Bickel[20], Balazs Banfai[20], Nathan P. Boley[20], James B. Brown[20], Haiyan Huang[20], Qunhua Li[20]†, Jingyi Jessica Li[20], William Stafford Noble[16,77], Jeffrey A. Bilmes[78], Orion J. Buske[16], Michael M. Hoffman[16], Avinash D. Sahu[16]†, Peter V. Kharchenko[79], Peter J. Park[79], Dannon Baker[80], James Taylor[80], Zhiping Weng[23], Sowmya Iyer[27], Xianjun Dong[23], Melissa Greven[23], Xinying Lin[23], Jie Wang[23], Hualin S. Xi[32], Jiali Zhuang[23], Mark Gerstein[21], Roger P. Alexander[21], Suganthi Balasubramanian[21], Chao Cheng[21], Arif Harmanci[21], Lucas Lochovsky[21], Renqiang Min[21]†, Xinmeng J. Mu[21], Joel Rozowsky[21], Koon-Kiu Yan[21], Kevin Y. Yip[21]† & Ewan Birney[1]

[1]Vertebrate Genomics Group, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK. [2]Department of Computer Science, Stanford University, 318 Campus Drive, Stanford, California 94305-5428, USA. [3]SwitchGear Genomics, 1455 Adams Drive Suite 1317, Menlo Park, California 94025, USA. [4]Functional Genomics, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. [5]College of Nanoscale Sciences and Engineering, University ay Albany-SUNY, 257 Fuller Road, NFE 4405, Albany, New York 12203, USA. [6]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. [7]Biochemistry and Molecular Biology, USC/Norris Comprehensive Cancer Center, 1450 Biggy Street, NRT 6503, Los Angeles, California 90089, USA. [8]Informatics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK. [9]Department of Medicine, Division of Medical Genetics, University of Washington, 3720 15th Avenue NE, Seattle, Washington 98195, USA. [10]College of Arts and Sciences, Boise State University, 1910 University Drive, Boise, Idaho 83725, USA. [11]Program in Systems Biology, Program in Gene Function and Expression, Department of Biochemistry and Molecular

Pharmacology, University of Massachusetts Medical School, 364 Plantation Street, Worcester, Massachusetts 01605, USA. [12]Department of Genetics, Stanford University, 300 Pasteur Drive, M-344, Stanford, California 94305-5120, USA. [13]Center for Systems and Synthetic Biology, Institute for Cellular and Molecular Biology, Section of Molecular Genetics and Microbiology, The University of Texas at Austin, 1 University Station A4800, Austin, Texas 78712, USA. [14]HudsonAlpha Institute for Biotechnology, 601 Genome Way, Huntsville, Alabama 35806, USA. [15]Center for Biomolecular Science and Engineering, University of California, Santa Cruz, 1156 High Street, Santa Cruz, California 95064, USA. [16]Department of Genome Sciences, University of Washington, 3720 15th Ave NE, Seattle, Washington 98195-5065, USA. [17]Institute for Genome Sciences and Policy, Duke University, 101 Science Drive, Durham, North Carolina 27708, USA. [18]Department of Biology, Carolina Center for Genome Sciences, and Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill, 408 Fordham Hall, Chapel Hill, North Carolina 27599-3280, USA. [19]Computer Science and Artificial Intelligence Laboratory, Broad Institute of MIT and Harvard, Massachusetts Institute of Technology, 32 Vassar Street, Cambridge, Massachusetts 02139, USA. [20]Department of Statistics, University of California, Berkeley, 367 Evans Hall, University of California, Berkeley, Berkeley, California 94720, USA. [21]Computational Biology and Bioinformatics Program, Yale University, 266 Whitney Avenue, New Haven, Connecticut 06520, USA. [22]Bioinformatics and Genomics, Centre for Genomic Regulation (CRG) and UPF, Doctor Aiguader, 88, Barcelona 08003, Catalonia, Spain. [23]Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, 364 Plantation Street, Worcester, Massachusetts 01605, USA. [24]Department of Genetics, The University of North Carolina at Chapel Hill, 120 Mason Farm Road, CB 7240, Chapel Hill, North Carolina 27599, USA. [25]Center for Comparative Genomics and Bioinformatics, The Pennsylvania State University, Wartik Laboratory, University Park, Pennsylvania 16802, USA. [26]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, 304 Wartik Laboratory, University Park, Pennsylvania 16802, USA. [27]Program in Bioinformatics, Boston University, 24 Cummington Street, Boston, Massachusetts 02215, USA. [28]RIKEN Omics Science Center, RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. [29]Division of Biology, California Institute of Technology, 156-291200 East California Boulevard, Pasadena, California 91125, USA. [30]Developmental and Cell Biology and Center for Complex Biological Systems, University of California Irvine, 2218 Biological Sciences III, Irvine, California 92697-2300, USA. [31]Genome Technology Branch, National Human Genome Research Institute, 5625 Fishers Lane, Bethesda, Maryland 20892, USA. [32]Department of Biochemistry and Molecular Pharmacology, Bioinformatics Core, University of Massachusetts Medical School, 364 Plantation Street, Worcester, Massachusetts 01605, USA. [33]Howard Hughes Medical Institute and Department of Pathology, Massachusetts General Hospital and Harvard Medical School, 185 Cambridge St CPZN 8400, Boston, Massachusetts 02114, USA. [34]National Human Genome Research Institute, National Institutes of Health, 31 Center Drive, Building 31, Room 4B09, Bethesda, Maryland 20892-2152, USA. [35]National Human Genome Research Institute, National Institutes of Health, 5635 Fishers Lane, Bethesda, Maryland 20892-9307, USA. [36]Department of Pediatrics, Division of Medical Genetics, Duke University School of Medicine, Durham, North Carolina 27710, USA. [37]National Human Genome Research Institute, National Institutes of Health, 5625 Fishers Lane, Rockville, Maryland 20892, USA. [38]Affymetrix, Inc., 3380 Central Expressway, Santa Clara, California 95051, USA. [39]Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Catalonia 08002, Spain. [40]Department of Genome Sciences, Box 355065, and Department of Medicine, Division of Oncology, Box 358081, University of Washington, Seattle, Washington 98195-5065, USA. [41]Institute for Genomics and Systems Biology, The University of Chicago, 900 East 57th Street, 10100 KCBD, Chicago, Illinois 60637, USA. [42]Beckman Institute, California Institute of Technology, 156-29 1200 E. California Boulevard, Pasadena, California 91125, USA. [43]Department of Biochemistry and Biophysics, University of North Carolina School of Medicine, Campus Box 7260, 120 Mason Farm Road, 3010 Genetic Medicine Building, Chapel Hill, North Carolina 27599, USA. [44]Centro Nacional de Análisis Genómico (CNAG), C/Baldiri Reixac 4, Torre I, Barcelona, Catalonia 08028, Spain. [45]Genomics, Affymetrix, Inc., 3380 Central Expressway, Santa Clara, California 95051, USA. [46]Center for Integrative Genomics, University of Lausanne, Genopode Building, 1015 Lausanne, Switzerland. [47]Genome Technology and Biology, Genome Institute of Singapore, 60 Biopolis Street, 02-01, Genome, Singapore 138672, Singapore. [48]Computational and Systems Biology, Genome Institute of Singapore, 60 Biopolis Street, 02-01, Genome, Singapore 138672, Singapore. [49]Department of Genetic Medicine and Development, University of Geneva Medical School, and University Hospitals of Geneva, 1 rue Michel-Servet, 1211 Geneva 4, Switzerland. [50]Department of Genetics, The University of North Carolina at Chapel Hill, 5078 GMB, Chapel Hill, North Carolina 27599-7264, USA. [51]Department of Biostatistics, Gillings School of Global Public Health, The University of North Carolina at Chapel Hill, 408 Fordham Hall, Chapel Hill, North Carolina 27599-7445, USA. [52]Center for Advanced Computing Research, California Institute of Technology, MC 158-79, 1200 East California Boulavard, Pasadena, California 91125, USA. [53]Department of Statistics, Stanford University, Sequoia Hall. 390 Serra Mall, Stanford, California 94305-4065, USA. [54]DOE Joint Genome Institute, Walnut Creek, California, USA. [55]Genomics Division, Lawrence Berkeley National Laboratory, One Cyclotron Road, MS 84-171, Berkeley, California 94720, USA. [56]Structural Computational Biology, Spanish National Cancer Research Centre (CNIO), Melchor Fernandez Almagro, 3, 28029 Madrid, Spain. [57]School of Life Sciences, Tsinghua University, School of Life Sciences, Tsinghua University, 100084 Beijing, China. [58]Department of Pathology and Laboratory Medicine, Institute for Computational Biomedicine, Weill Cornell Medical College, 1305 York Avenue, Box 140, New York, New York 10065, USA. [59]Computer Science and Engineering, Washington University in St Louis, St Louis, Missouri 63130, USA. [60]Department of Genetics, Albert Einstein College of Medicine, 1301 Morris Park Avenue, Room 353A, Bronx, New York 10461, USA. [61]Center for Biomolecular Science and Engineering, Howard Hughes Medical Institute, University of California, Santa Cruz, 1156 High Street, Santa Cruz, California 95064, USA. [62]Genome Center, University of California-Davis, 451 Health Sciences Drive, Davis, California 95616, USA. [63]Department of Molecular, Cellular, and Developmental Biology, Yale University, 266 Whitney Avenue, New Haven, Connecticut 06511, USA. [64]Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA. [65]Biochemistry and Molecular Biology, University of Southern California, 1501 San Pablo Street, Los Angeles, California 90089, USA. [66]Department of Biomedical Informatics, Ohio State University, 3172C Graves Hall, 333 W Tenth Avenue, Columbus, Ohio 43210, USA. [67]Department of Genetics, Yale University, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA. [68]Department of Cellular and Structural Biology, Children's Cancer Research Institute–UTHSCSA, Mail code 7784- 7703 Floyd Curl Dr, San Antonio, Texas 78229, USA. [69]Centre for Organismal Studies (COS) Heidelberg, University of Heidelberg, Im Neuenheimer Feld 230, 69120 Heidelberg, Germany. [70]Basic Sciences Division, Fred Hutchinson Cancer Research Center, 825 Eastlake Avenue East, Seattle, Washington 98109, USA. [71]Department of Medicine, Division of Medical Genetics, Box 357720, University of Washington, Seattle, Washington 98195-7720, USA. [72]Division of Human Biology, Fred Hutchinson Cancer Research Center, 825 Eastlake Avenue East, Seattle, Washington 98109, USA. [73]Department of Psychiatry and Behavioral Sciences, Box 356560, University of Washington, Seattle, Washington 98195-6560, USA. [74]Microarray Informatics Group, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK. [75]Genomics and Regulatory Systems Group, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK. [76]Department of Pathology, Department of Genetics, Stanford University, 300 Pasteur Drive, Stanford, California 94305, USA. [77]Department of Computer Science and Engineering, 185 Stevens Way, Seattle, Washington 98195, USA. [78]Department of Electrical Engineering, University of Washington, 185 Stevens Way, Seattle, Washington 98195, USA. [79]Center for Biomedical Informatics, Harvard Medical School, 10 Shattuck Street, Boston, Massachusetts 02115, USA. [80]Departments of Biology and Mathematics and Computer Science, Emory University, Atlanta, Georgia 30322, USA. †Present addresses: Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 32 Vassar Street, Cambridge, Massachusetts 02139, USA (A.K.); UCLA Biological Chemistry Department, Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research at UCLA, Jonsson Comprehensive Cancer Center, 615 Charles E Young Dr South, Los Angeles, California 90095, USA (J.E.); Department of Statistics, 514D Wartik Lab, Penn State University, State College, Pennsylvania 16802, USA (Q.L.); Department of Biostatistics and Bioinformatics and the Institute for Genome Sciences and Policy, Duke University School of Medicine, 101 Science Drive, Durham, North Carolina 27708, USA (T.E.R.); Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong (K.Y.Y.); Department of Genetics, Washington University in St Louis, St Louis, Missouri 63110, USA (R.F.L.); Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, Maryland 20742, USA (L.A.L.D.); National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA (J.Z.); University of California, Davis Population Biology Graduate Group, Davis, California 95616, USA (J.R.W.); Illumina Cambridge Ltd., Chesterford Research Park, Little Chesterford, Saffron Walden, Essex CB10 1XL, UK (E.H.M.); BlueGnome Ltd., CPC4, Capital Park, Fulbourn, Cambridge CB21 5XE, UK (F.K.); Institut de Génétique et Développement de Rennes, CNRS-UMR6061, Université de Rennes 1, F-35000 Rennes, Brittany, France (T.D.); Caltech, 1200 East California Boulevard, Pasadena, California 91125, USA (K.F.-T.); A*STAR-Duke-NUS Neuroscience Research Partnership, 8 College Road, Singapore 169857, Singapore (M.J.F.); St Laurent Institute, One Kendall Square, Cambridge, Massachusetts 02139, USA (P.K.); Department of Genetics, Stanford University, Stanford, California 94305, USA (H.T.); Biomedical Sciences (BMS) Graduate Program, University of California, San Francisco, 513 Parnassus Avenue, HSE-1285, San Francisco, California 94143-0505, USA (S.L.P.); Monterey Bay Aquarium Research Institute, Moss Landing, California 95039, USA (M.J.v.B.); Department of Machine Learning, NEC Laboratories America, 4 Independence Way, Princeton, New Jersey 08540, USA (R.M.); Neuronal Circuit Development Group, Unité de Génétique et Biologie du Développement, U934/UMR3215, Institut Curie–Centre de Recherche, Pole de Biologie du Développement et Cancer, 26, rue d'Ulm, 75248 Paris Cedex 05, France (T.A.); Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK (M.L.); Unidade de Bioinformatica, Rua da Quinta Grande, 6, P-2780-156 Oeiras, Portugal (D.S.); Department of Genome Sciences, University of Washington, 3720 15th Avenue NE, Seattle, Washington 98195-5065, USA (M.W.L.); Center for Bioinformatics and Computational Biology, 3115 Ag/Life Surge Building 296, University of Maryland, College Park, Maryland 20742, USA (A.D.S.).

ARTICLE

doi:10.1038/nature11232

# The accessible chromatin landscape of the human genome

Robert E. Thurman[1,*], Eric Rynes[1,*], Richard Humbert[1,*], Jeff Vierstra[1], Matthew T. Maurano[1], Eric Haugen[1], Nathan C. Sheffield[2], Andrew B. Stergachis[1], Hao Wang[1], Benjamin Vernot[1], Kavita Garg[3], Sam John[1], Richard Sandstrom[1], Daniel Bates[1], Lisa Boatman[4], Theresa K. Canfield[1], Morgan Diegel[1], Douglas Dunn[1], Abigail K. Ebersol[4], Tristan Frum[4], Erika Giste[1], Audra K. Johnson[1], Ericka M. Johnson[4], Tanya Kutyavin[1], Bryan Lajoie[5], Bum-Kyu Lee[6], Kristen Lee[1], Darin London[2], Dimitra Lotakis[4], Shane Neph[1], Fidencio Neri[1], Eric D. Nguyen[4], Hongzhu Qu[1,7], Alex P. Reynolds[1], Vaughn Roach[1], Alexias Safi[2], Minerva E. Sanchez[4], Amartya Sanyal[5], Anthony Shafer[1], Jeremy M. Simon[8], Lingyun Song[2], Shinny Vong[1], Molly Weaver[1], Yongqi Yan[4], Zhancheng Zhang[8], Zhuzhu Zhang[8], Boris Lenhard[9,†], Muneesh Tewari[3], Michael O. Dorschner[10], R. Scott Hansen[4], Patrick A. Navas[4], George Stamatoyannopoulos[4], Vishwanath R. Iyer[6], Jason D. Lieb[8], Shamil R. Sunyaev[11], Joshua M. Akey[1], Peter J. Sabo[1], Rajinder Kaul[4], Terrence S. Furey[8], Job Dekker[5], Gregory E. Crawford[2] & John A. Stamatoyannopoulos[1,12]

**DNase I hypersensitive sites (DHSs) are markers of regulatory DNA and have underpinned the discovery of all classes of *cis*-regulatory elements including enhancers, promoters, insulators, silencers and locus control regions. Here we present the first extensive map of human DHSs identified through genome-wide profiling in 125 diverse cell and tissue types. We identify ~2.9 million DHSs that encompass virtually all known experimentally validated *cis*-regulatory sequences and expose a vast trove of novel elements, most with highly cell-selective regulation. Annotating these elements using ENCODE data reveals novel relationships between chromatin accessibility, transcription, DNA methylation and regulatory factor occupancy patterns. We connect ~580,000 distal DHSs with their target promoters, revealing systematic pairing of different classes of distal DHSs and specific promoter types. Patterning of chromatin accessibility at many regulatory regions is organized with dozens to hundreds of co-activated elements, and the transcellular DNase I sensitivity pattern at a given region can predict cell-type-specific functional behaviours. The DHS landscape shows signatures of recent functional evolutionary constraint. However, the DHS compartment in pluripotent and immortalized cells exhibits higher mutation rates than that in highly differentiated cells, exposing an unexpected link between chromatin accessibility, proliferative potential and patterns of human variation.**

**ENCODE**
Encyclopedia of DNA Elements
nature.com/encode

Cell-selective activation of regulatory DNA drives the gene expression patterns that shape cell identity. Regulatory DNA is characterized by the cooperative binding of sequence-specific transcriptional regulatory factors in place of a canonical nucleosome, leading to a remodelled chromatin state characterized by markedly heightened accessibility to nucleases[1]. DNase I hypersensitive sites (DHSs) in chromatin were first identified over 30 years ago, and have since been used extensively to map regulatory DNA regions in diverse organisms[2]. DNase I hypersensitivity is central to all defined classes of active *cis*-regulatory elements including enhancers, promoters, silencers, insulators and locus control regions[2–4]. Because DNase I hypersensitivity overlies *cis*-regulatory elements directly and is maximal over the core region of regulatory factor occupancy, it enables precise delineation of the genomic *cis*-regulatory compartment. DHSs are flanked by nucleosomes, which may acquire histone modification patterns that reflect the functional role of the adjoining regulatory DNA, such as the association of histone H3 lysine 4 trimethylation (H3K4me3) with promoter elements[5]. Recent advances have enabled genome-scale mapping of DHSs in mammalian cells[6–8], laying the foundations for comprehensive catalogues of human regulatory DNA.

## General features of the accessible chromatin landscape

Two ENCODE production centres (University of Washington and Duke University) profiled DNase I sensitivity genome-wide using massively parallel sequencing[7–9] in a total of 125 human cell and tissue types including normal differentiated primary cells ($n = 71$), immortalized primary cells ($n = 16$), malignancy-derived cell lines ($n = 30$) and multipotent and pluripotent progenitor cells ($n = 8$) (Supplementary Table 1). The density of mapped DNase I cleavages as a function of genome position provides a continuous quantitative measure of chromatin accessibility, in which DHSs appear as prominent peaks within the signal data from each cell type (Fig. 1a and Supplementary Figs 1 and 2). Analysis using a common algorithm (see Methods) identified 2,890,742 distinct high-confidence DHSs (false discovery rate (FDR) of 1%; see Methods), each of which was active in one or more cell types. Of these DHSs, 970,100 were specific to a single cell type, 1,920,642 were active in 2 or more cell types, and a

---

[1]Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. [2]Institute for Genome Sciences and Policy, Duke University, Durham, North Carolina 27708, USA. [3]Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. [4]Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, Washington 98195, USA. [5]Program in Systems Biology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA. [6]Institute for Cellular and Molecular Biology, University of Texas, Austin, Texas 78712, USA. [7]Laboratory of Disease Genomics and Individualized Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100029, China. [8]Department of Biology, University of North Carolina, Chapel Hill, North Carolina 27599, USA. [9]Department of Biology and Bergen Center for Computational Science, University of Bergen, Bergen 5008, Norway. [10]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Washington 98195, USA. [11]Department of Medicine, Division of Genetics, Brigham & Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. [12]Department of Medicine, Division of Oncology, University of Washington, Seattle, Washington 98195, USA. †Present address: Institute for Clinical Sciences, Faculty of Medicine, Imperial College London, and MRC Clinical Sciences Centre, London W12 0NN, UK.
*These authors contributed equally to this work.

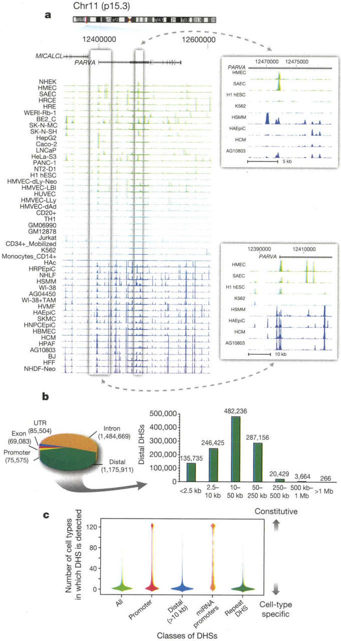


**Figure 1 | General features of the DHS landscape.** **a**, Density of DNase I cleavage sites for selected cell types, shown for an example ~350-kb region. Two regions are shown to the right in greater detail. **b**, Left: distribution of 2,890,742 DHSs with respect to GENCODE gene annotations. Promoter DHSs are defined as the first DHS localizing within 1 kb upstream of a GENCODE TSS. Right: distribution of intergenic DHSs relative to Gencode TSSs. **c**, Distributions of the number of cell types, from 1 to 125 ($y$ axis), in which DHSs in each of four classes ($x$ axis) are observed. Width of each shape at a given $y$ value shows the relative frequency of DHSs present in that number of cell types.

small minority (3,692) was detected in all cell types. The relative accessibility of DHSs along the genome varies by >100-fold and is highly consistent across cell types (Supplementary Figs 1 and 2). To estimate the sensitivity and accuracy of the sequencing-derived DHS maps, one ENCODE production centre (University of Washington) performed 7,478 classical DNase I hypersensitivity experiments by the Southern hybridization method[2]. Using Southern blots as the standard, the average sensitivity, per cell type, of DNase I-seq (at a sequencing depth of 30 M uniquely mapping reads) was 81.6%, with specificity of 99.5–99.9%. Of DHSs classified as false negatives within a particular cell type, an average of 92.4% were detected as a DHS in another cell type or upon deeper sequencing. As such, we estimate that the overall sensitivity for DHSs of the combined cell type maps is >98%.

Approximately 3% ($n = 75{,}575$) of DHSs localize to transcriptional start sites (TSSs) defined by GENCODE[10] and 5% ($n = 135{,}735$, including the aforementioned) lie within 2.5 kilobases (kb) of a TSS. The remaining 95% of DHSs are positioned more distally, and are roughly evenly divided between intronic and intergenic regions (Fig. 1b). Promoters typically exhibit high accessibility across cell types, with the average promoter DHS detected in 29 cell types (Fig. 1c, second column). By contrast, distal DHSs are largely cell selective (Fig. 1c, third column).

MicroRNAs (miRNAs) comprise a major class of regulatory molecules and have been extensively studied, resulting in consensus annotation of hundreds of conserved miRNA genes[11], approximately one-third of which are organized in polycistronic clusters[12]. However, most predicted promoters driving microRNA expression lack experimental evidence. Of 329 unique annotated miRNA TSSs (Supplementary Methods), 300 (91%) either coincided with or closely approximated (<500 base pairs (bp)) a DHS. Chromatin accessibility at miRNA promoters was highly promiscuous compared with GENCODE TSSs (Fig. 1c, fourth column), and showed cell lineage organization, paralleling the known regulatory roles of well-annotated lineage-specific miRNAs (Supplementary Fig. 3).

The 20–50-bp read lengths from DNase I-seq experiments enabled unique mapping to 86.9% of the genomic sequence, allowing us to interrogate a large fraction of transposon sequences. A surprising number contain highly regulated DHSs (Fig. 1c, fifth column and Supplementary Figs 4 and 5), compatible with cell-specific transcription of repetitive elements detected using ENCODE RNA sequencing data[13]. DHSs were most strongly enriched at long terminal repeat (LTR) elements, which encode retroviral enhancer structures (Supplementary Table 2). Two such examples are shown in Supplementary Fig. 4, which also illustrates the strong cell-selectivity of chromatin accessibility seen for each major repeat class. We also documented numerous examples of transposon DHSs that displayed enhancer activity in transient transfection assays (Supplementary Table 3).

Comparison with an extensive compilation of 1,046 experimentally validated distal, non-promoter *cis*-regulatory elements (enhancers, insulators, locus control regions, and so on) revealed the overwhelming majority (97.4%) to be encompassed within DNase I hypersensitive chromatin (Supplementary Table 4), typically with strong cell selectivity (Supplementary Fig. 2b).

## Transcription factor drivers of chromatin accessibility

DNase I hypersensitive sites result from cooperative binding of transcriptional factors in place of a canonical nucleosome[1,2]. To quantify the relationship between chromatin accessibility and the occupancy of regulatory factors, we compared sequencing-depth-normalized DNase I sensitivity in the ENCODE common cell line K562 to normalized chromatin immunoprecipitation and high-throughput sequencing (ChIP-seq) signals from all 42 transcription factors mapped by ENCODE ChIP-seq[14] in this cell type (Fig. 2). Simple summation of the ChIP-seq signals markedly parallels quantitative DNase I sensitivity at individual DHSs (Fig. 2a) and across the genome ($r = 0.79$, Fig. 2b). For example, the β-globin locus control region contains a major enhancer element at hypersensitive site 2 (HS2), which appears to be occupied by dozens of transcription factors (Supplementary Fig. 6a). Such highly overlapping binding patterns have been interpreted to signify weak interactions with lower-affinity recognition sequences potentiated by an accessible DNA template[15]. However, HS2 is a compact element with a functional core spanning ~110 bp that contains 5–8 sites of transcription factor–DNA interaction *in vivo* depending on the cell type[16–18]. The fact that the cumulative ChIP-seq signal closely parallels the degree of nuclease sensitivity at HS2 and elsewhere is thus most readily explained by interactions between DNA-bound factors

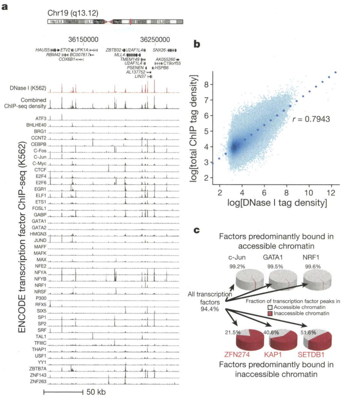


**Figure 2 | Transcription factor drivers of chromatin accessibility. a**, DNase I tag density is shown in red for a 175-kb region of chromosome 19. Below: normalized ChIP-seq tag density for 45 ENCODE ChIP-seq experiments from K562 cells, with a cumulative sum of the individual tag density tracks shown immediately below the K562 DNase I data. **b**, Genome-wide correlation ($r = 0.7943$) between ChIP-seq and DNase I tag densities ($\log_{10}$) in K562 cells. **c**, Left: 94.4% of a combined 1,108,081 ChIP-seq peaks from all transcription factors assayed in K562 cells fall within accessible chromatin (grey areas of pie chart). Top: three examples of transcription factors localizing almost exclusively within accessible chromatin. Bottom: three transcription factors from the KRAB-associated complex localizing partially or predominantly within inaccessible chromatin.

and other interacting factors that collectively potentiate the accessible chromatin state (Supplementary Fig. 6b). Given the relatively limited number of factors studied, it may seem surprising that such a close correlation should be evident. However, most of the factors selected for ENCODE ChIP-seq studies have well-described or even fundamental roles in transcriptional regulation, and many were identified originally based on their high affinity for DNA. Alternatively, as originally proposed in ref. 19, a limited number of factors may be involved in establishment and maintenance of chromatin remodelling, whereas others may interact nonspecifically with the remodelled state. We also found that the recognition sequences for a small number of factors were consistently linked with elevated chromatin accessibility across all classes of sites and all cell types (Supplementary Fig. 6c), indicating that regulators acting through these sequences are key drivers of the accessibility landscape.

Overall, 94.4% of a combined 1,108,081 ChIP-seq peaks from all ENCODE transcription factors fall within accessible chromatin (Fig. 2c and Supplementary Fig. 7a), with the median factor having 98.2% of its binding sites localized therein. Notably, a small number of factors diverged from this paradigm, including known chromatin repressors, such as the KRAB-associated factors KAP1 (also called TRIM28), SETDB1 and ZNF274 (refs 20, 21) (Fig. 2c). We hypothesized that a proportion of the occupancy sites of these factors represented binding within compacted heterochromatin. To test this, we developed targeted mass spectrometry assays[22] for KAP1 and three factors localizing almost exclusively within accessible chromatin (GATA1, c-Jun, NRF1), and quantified their abundance in biochemically defined heterochromatin[23] against a total chromatin fraction (Supplementary Fig. 7b). This analysis confirmed that factors such as KAP1 show a significant level of heterochromatin occupancy (Supplementary Fig. 7c).

## An invariant directional promoter chromatin signature

The annotation of sites of transcription origination continues to be an active and fundamental endeavour[13]. In addition to direct evidence of TSSs provided by RNA transcripts, H3K4me3 modifications are closely linked with TSSs[24]. We therefore explored systematically the relationship between chromatin accessibility and H3K4me3 patterns at well-annotated promoters, its relationship to transcription origination, and its variability across ENCODE cell types.

We performed ChIP-seq for H3K4me3 in 56 cell types using the same biological samples used for DNase I data (Supplementary Table 1, column D). Plotting DNase I cleavage density against ChIP-seq tag density around TSSs reveals highly stereotyped, asymmetrical patterning of these chromatin features with a precise relationship to the TSS (Fig. 3a, b). This directional pattern is consistent with a rigidly positioned nucleosome immediately downstream from the promoter DHS, and is largely invariant across cell types (Fig. 3b and Supplementary Fig. 8).

To map novel promoters (and their directionality) not encompassed by the GENCODE consensus annotations, we applied a pattern-matching approach to scan the genome across all 56 cell types (Supplementary Methods). Using this approach we identified a total of 113,622 distinct putative promoters. Of these, 68,769 correspond to previously annotated TSSs, and 44,853 represent novel predictions (versus GENCODE v7). Of the novel sites, 99.5% are supported by evidence from spliced expressed sequence tags (ESTs) and/or cap analysis of gene expression (CAGE) tag clusters (Fig. 3c and Supplementary Fig. 9, $P < 0.0001$; see Supplementary Methods). We found novel sites in every configuration relative to existing annotations (Fig. 3d–f and Supplementary Fig. 10). For example, 29,203 putative promoters are contained in the bodies of annotated genes, of which 17,214 are oriented antisense to the annotated direction of transcription, and 2,794 lie immediately downstream of an annotated gene's 3′ end, with 1,638 in antisense orientation. The results indicate that chromatin data can systematically inform RNA transcription analyses, and suggest the existence of a large pool of cell-selective transcriptional promoters, many of which lie in antisense orientations.

## Chromatin accessibility and DNA methylation patterns

CpG methylation has been closely linked with gene regulation, based chiefly on its association with transcriptional silencing[25]. However, the relationship between DNA methylation and chromatin structure has not been clearly defined. We analysed ENCODE reduced-representation bisulphite sequencing (RRBS) data, which provide quantitative methylation measurements for several million CpGs (K. E. Varley *et al.*, manuscript submitted; see Gene Expression Omnibus accession GSE27584). We focused on 243,037 CpGs falling within DHSs in 19 cell types for which both data types were available from the same sample. We observed two broad classes of sites: those with a strong inverse correlation across cell types between DNA methylation and chromatin accessibility (Fig. 4a and Supplementary Fig. 11a), and those with variable chromatin accessibility but constitutive hypomethylation (Fig. 4a, right). To quantify these trends globally, we performed a linear regression analysis between chromatin accessibility and DNA methylation at the 34,376 CpG-containing DHSs (see Supplementary Methods). Of these sites, 6,987 (20%) showed a significant association (1% FDR) between methylation and accessibility (Supplementary Fig. 11b). Increased methylation was almost uniformly negatively associated with chromatin accessibility (>97% of cases). The magnitude of the association between methylation and accessibility was strong, with the latter on

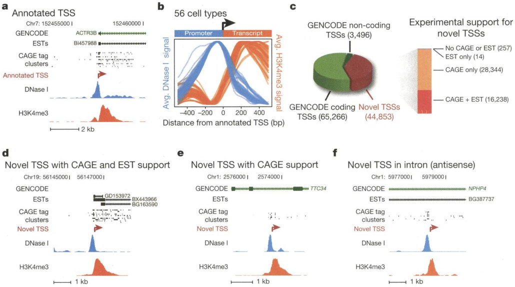


**Figure 3 | Identification and directional classification of novel promoters.** **a**, DNase I (blue) and H3K4me3 (red) tag densities for K562 cells around annotated TSS of *ACTR3B*. **b**, Averaged H3K4me3 tag density (red, right *y* axis) and log DNase I tag density (blue, left *y* axis) across 10,000 randomly selected GENCODE TSSs, oriented 5′→3′. Each blue and red curve is for a different cell type, showing invariance of the pattern. **c**, Relation of 113,615 promoter predictions to GENCODE annotations, with supporting EST and CAGE evidence (bar at right). **d–f**, Examples of novel promoters identified in K562; red arrow marks predicted TSS and direction of transcription, with CAGE tag clusters, spliced ESTs and GENCODE annotations above. **d**, Novel TSS confirmed by CAGE and ESTs. **e**, Novel TSS confirmed by CAGE, no ESTs. Note intronic location. **f**, Antisense prediction within annotated gene.

average 95% lower in cell types with coinciding methylation versus cell types lacking coinciding methylation (Supplementary Fig. 11c). Fully 40% of variable methylation was associated with a concomitant effect on accessibility.

The role of DNA methylation in causation of gene silencing is presently unclear. Does methylation reduce chromatin accessibility by evicting transcription factors? Or does DNA methylation passively 'fill in' the voids left by vacating transcription factors? Transcription factor expression is closely linked with the occupancy of its binding sites[26]. If the former of the two above hypotheses is correct, methylation of individual binding site sequences should be independent of transcription factor gene expression. If the latter, methylation at transcription factor recognition sequences should be negatively correlated with transcription factor abundance (Fig. 4b).

Comparing transcription factor transcript levels to average methylation at cognate recognition sites within DHSs revealed significant negative correlations between transcription factor expression and binding site methylation for most (70%) transcription factors with a significant association ($P < 0.05$). Representative examples are shown in Fig. 4c and Supplementary Fig. 12a. These data argue strongly that methylation patterning paralleling cell-selective chromatin accessibility results from passive deposition after the vacation of transcription factors from regulatory DNA, confirming and extending other recent reports[27].

Interestingly, a small number of factors showed positive correlations between expression and binding site methylation (Supplementary Fig. 12b), including MYB and LUN-1 (also known as TOPORS). Both of these transcription factors showed increased transcription

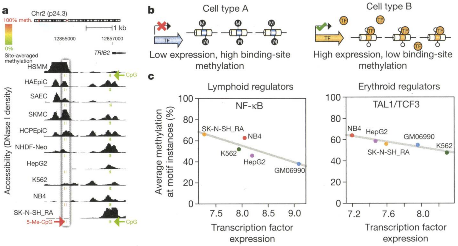


**Figure 4 | Chromatin accessibility and DNA methylation patterns.** **a**, DNase I sensitivity in 10 cell types with ENCODE reduced representation bisulphite sequencing data. Inset box: accessibility (*y* axis) decreases quantitatively as methylation increases. Other DHSs (right) show low correlation between accessibility and methylation. CpG methylation scale: green, 0%; yellow, 50%; red, 100%. **b**, Model of transcription factor (TF)-driven methylation patterns in which methylation passively mirrors transcription factor occupancy. **c**, Relationship between transcription factor transcript levels and overall methylation at cognate recognition sequences of the same transcription factors. Lymphoid regulators in B-lymphoblastoid line GM06990 (left) and erythroid regulators in the erythroleukaemia line K562 (right). Negative correlation indicates that site-specific DNA methylation follows transcription factor vacation of differentially expressed transcription factors.

and binding site methylation specifically within acute promyelocytic leukaemia cells (NB4), and both interact with promyelocytic leukaemia (PML) bodies[28,29], a sub-nuclear structure disrupted in PML cells. The anomalous behaviour of these two transcription factors with respect to chromatin structure and DNA methylation may thus be related to a specialized mechanism seen only in pathologically altered cells.

## A map of distal DHS-to-promoter connections

From examination of DNase I profiles across many cell types we observed that many known cell-selective enhancers become DHSs synchronously with the appearance of hypersensitivity at the promoter of their target gene (Supplementary Fig. 13). To generalize this, we analysed the patterning of 1,454,901 distal DHSs (DHSs separated from a TSS by at least one other DHS) across 79 diverse cell types (Supplementary Methods and Supplementary Table 6), and correlated the cross-cell-type DNase I signal at each DHS position with that at all promoters within ±500 kb (Supplementary Fig. 14a). We identified a total of 578,905 DHSs that were highly correlated ($r > 0.7$) with at least one promoter ($P < 10^{-100}$), providing an extensive map of candidate enhancers controlling specific genes (Supplementary Methods and Supplementary Table 7). To validate the distal DHS/enhancer–promoter connections, we profiled chromatin interactions using the chromosome conformation capture carbon copy (5C) technique[30]. For example, the phenylalanine hydroxylase (*PAH*) gene is expressed in hepatic cells, and an enhancer has been defined upstream of its TSS (Fig. 5a). The correlation values for three DHSs within the gene body closely parallel the frequency of long-range chromatin interactions measured by 5C. The three interacting intronic DHSs cloned downstream of a reporter gene driven by the *PAH* promoter all showed increased expression ranging from three- to tenfold over a promoter-only control, confirming enhancer function.

We next examined comprehensive promoter-versus-all 5C experiments performed over 1% of the human genome[31] in K562 cells. DHS–promoter pairings were markedly enriched in the specific cognate chromatin interaction ($P < 10^{-13}$, Supplementary Fig. 14b). We also examined K562 promoter–DHS interactions detected by polymerase II chromatin interaction analysis with paired-end tag sequencing (ChIA-PET)[24], which quantifies interactions between promoter-bound polymerase and distal sites. The ChIA-PET interactions were also markedly enriched for DHS–promoter pairings ($P < 10^{-15}$, Supplementary Fig. 14c). Together, the large-scale interaction analyses affirm the fidelity of DHS–promoter pairings based on correlated DNase I sensitivity signals at distal and promoter DHSs.

Most promoters were assigned to more than one distal DHS, indicating the existence of combinatorial distal regulatory inputs for most genes (Fig. 5b and Supplementary Table 7). A similar result is forthcoming from large-scale 5C interaction data[31]. Surprisingly, roughly half of the promoter-paired distal DHSs were assigned to more than one promoter (Fig. 5b and Supplementary Methods), indicating that human *cis*-regulatory circuitry is significantly more complicated than previously anticipated, and may serve to reinforce the robustness of cellular transcriptional programs.

The number of distal DHSs connected with a particular promoter provides, for the first time, a quantitative measure of the overall regulatory complexity of that gene. We asked whether there are any systematic functional features of genes with highly complex regulation. We ranked all human genes by the number of distal DHSs paired with the promoter of each gene, then performed a Gene Ontology analysis on the rank-ordered list. We found that the most complexly regulated human genes were markedly enriched in immune system functions (Supplementary Fig. 14d), indicating that the complexity of cellular and environmental signals processed by the immune system is directly encoded in the *cis*-regulatory architecture of its constituent genes.

Next, we asked whether DHS–promoter pairings reflected systematic relationships between specific combinations of regulatory factors (Supplementary Methods). For example, KLF4, SOX2, OCT4 (also called POU5F1) and NANOG are known to form a well-characterized transcriptional network controlling the pluripotent state of embryonic stem cells[32]. We found significant enrichment ($P < 0.05$) of the KLF4, SOX2 and OCT4 motifs within distal DHSs correlated with promoter DHSs containing the NANOG motif; enrichment of NANOG, SOX2 and OCT4 distal motifs co-occurring with promoter motif OCT4; and enrichment of distal SOX2 and OCT4 motifs with promoter SOX2 motifs (Supplementary Fig. 15a). By contrast, promoters containing KLF4 motifs were associated with KLF4-containing distal DHSs, but not with DHSs containing NANOG, SOX2 or OCT4 motifs (Supplementary Fig. 15a, bottom).

We also tested for significant co-associations between promoter types (defined by the presence of cognate motif classes; see Supplementary Methods) and motifs in paired distal DHSs (Fig. 5c and Supplementary Fig. 15b, c). For example, when a member of the ETS domain family (motifs ETS1, ETS2, ELF1, ELK1, NERF (also called ELF2), SPIB, and others) is present within a promoter DHS, motif PU.1 (also called SPI1) is significantly more likely to be observed in a correlated distal DHS ($P < 10^{-5}$). These results suggest that a limited set of general rules may govern the pairing of co-regulated distal DHSs with particular promoters.

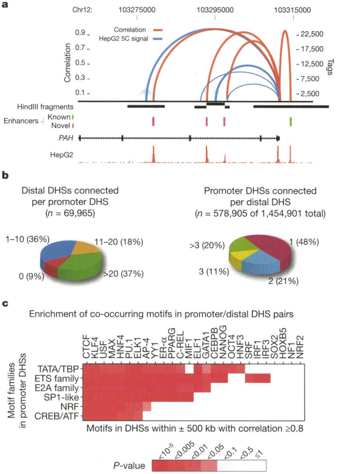


**Figure 5 | A genome-wide map of distal DHS-to-promoter connectivity.** **a**, Cross-cell-type correlation (red arcs, left *y* axis) of distal DHSs and *PAH* promoter closely parallels chromatin interactions measured by 5C-seq (blue arcs, right *y* axis); black bars indicate HindIII fragments used in 5C assays. Known (green) and novel (magenta) enhancers confirmed in transfection assays are shown below. Enhancer at far right is not separable by 5C as it lies within the HindIII fragment containing the promoter. **b**, Left: proportions of 69,965 promoters correlated ($r > 0.7$) with 0 to >20 DHSs within 500 kb. Right: proportions of 578,905 non-promoter DHSs (out of 1,454,901) correlated with 1 to >3 promoters within 500 kb. **c**, Pairing of canonical promoter motif families with specific motifs in distal DHSs.

## Stereotyped chromatin accessibility parallels function

In addition to the synchronized activation of distal DHSs and promoters described above, we observed a surprising degree of patterned co-activation among distal DHSs, with nearly identical cross-cell-type patterns of chromatin accessibility at groups of DHSs widely separated *in trans* (Supplementary Figs 16 and 17). For many patterns, we observed tens or even hundreds of like elements around the genome. The simplest explanation is that such co-activated sites share recognition motifs for the same set of regulatory factors. We found, however, that the underlying sequence features for a given pattern were surprisingly plastic. This suggests that the same pattern of cell-selective chromatin accessibility shared between two DHSs can be achieved by distinct mechanisms, probably involving complex combinatorial tuning.

We next asked whether distal DHSs with specific functions such as enhancers exhibited stereotypical patterning, and whether such patterning could highlight other elements with the same function. We examined one of the best-characterized human enhancers, DNase I HS2 of the β-globin locus control region[16–18]. HS2 is detected in many cell types, but exhibits potent enhancer activity only in erythroid cells[33]. Using a pattern-matching algorithm (see Supplementary Methods) we identified additional DHSs with nearly identical cross-cell-type accessibility patterns (Fig. 6a). We selected 20 elements across the spectrum of the top 200 matches to the HS2 pattern, and tested these in transient transfection assays in K562 cells (Supplementary Methods). Seventy per cent (14 of 20) of these displayed enhancer activity (mean 8.4-fold over control) (Fig. 6a, f). Of note, one (E3) showed a greater magnitude of enhancement (18-fold versus control) than HS2, which is itself one of the most potent known enhancers[4]. Next we selected three elements from the 14 HS2-like enhancers, applied pattern matching (Methods) to each to identify stereotyped elements, and tested samples of each pattern for enhancer activity, revealing additional K562 enhancers (total 15 of 25 positive) (Fig. 6b–d, f). In each case, therefore, we were able to discover enhancers by simply anchoring on the cross-cell-type DHS pattern of an element with enhancer activity. Collectively, these results show that co-activation of DHSs reflected in cross-cell-type patterning of chromatin accessibility is predictive of functional activity within a specific cell type, and suggest more generally that DHSs with stereotyped cellular patterning are likely to fulfil similar functions.

To visualize the qualities and prevalence of different stereotyped cross-cellular DHS patterns, we constructed a self-organizing map of a random 10% subsample of DHSs across all cell types and identified a total of 1,225 distinct stereotyped DHS patterns (Supplementary Figs 18 and 19). Many of the stereotyped patterns discovered by the self-organizing map encompass large numbers of DHSs, with some counting >1,000 elements (Supplementary Fig. 20).

Taken together, the above results show that chromatin accessibility at regulatory DNA is highly choreographed across large sets of co-activated elements distributed throughout the genome, and that DHSs with similar cross-cell-type activation profiles probably share similar functions.

## Variation in regulatory DNA linked to mutation rate

The DHS compartment as a whole is under evolutionary constraint, which varies between different classes and locations of elements[14], and may be heterogeneous within individual elements[34]. To understand the evolutionary forces shaping regulatory DNA sequences in humans, we estimated nucleotide diversity ($\pi$) in DHSs using publicly available whole-genome sequencing data from 53 unrelated individuals[35] (see Supplementary Methods). We restricted our analysis to nucleotides outside of exons and RepeatMasked regions. To provide a comparison with putatively neutral sites, we computed $\pi$ in fourfold degenerate synonymous positions (third positions) of coding exons. This analysis showed that, taken together, DHSs exhibit lower $\pi$ than fourfold degenerate sites, compatible with the action of purifying selection.

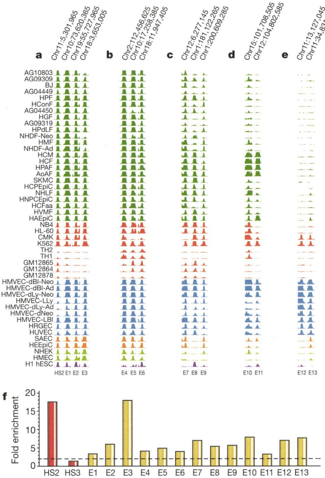


**Figure 6 | Stereotyped regulation of chromatin accessibility. a–e,** Enhancers grouped by similar chromatin stereotypes. Related cell lines are colour matched. HS2 from the β-globin locus control region is at left. E1–E11 represent progressively weaker matches to the HS2 stereotype. E12–13 derive from matches to a different stereotype based on another K562 enhancer. **f**, Experimental validation of enhancers detected by pattern matching. Bars indicate fold enrichment observed in transient assays in K562 relative to promoter-only control; mean of testing in both orientations is shown. Red bars indicate data from two potent *in vivo* enhancers, β-globin LCR HS2 and HS3; the latter requires chromatinization to function and is not active in transient assays. Gold bars indicate data from E1–E13 from **a–e** above.

Figure 7a shows $\pi$ for the DHSs of all analysed cell types, with colour coding to indicate the origin of each cell type. Particularly striking is the distribution of diversity relative to proliferative potential. DHSs in cells with limited proliferative potential have uniformly lower average diversity than immortal cells, with the difference most pronounced in malignant and pluripotent lines. This ordering is identical when highly mutable CpG nucleotides are removed from the analysis.

If differences in $\pi$ are due to mutation rate differences in different DHS compartments, the ratio of human polymorphism to human–chimpanzee divergence should remain constant across cell types. By contrast, differences in $\pi$ due to selective constraint should result in pronounced differences. To distinguish between these alternatives, we first compared polymorphism and human–chimpanzee divergence for DHSs from normal, malignant and pluripotent cells (Fig. 7b). Differences in polymorphism and divergence between these three groups are nearly identical, compatible with a mutational cause. Second, raw mutation rate is expected to affect rare and common genetic variation equally, whereas selection is likely to have a larger impact on common variation. We consistently observe ~62% of single nucleotide polymorphisms (SNPs) in DHSs of each group to have derived-allele frequencies below 0.05. DHSs in different cell

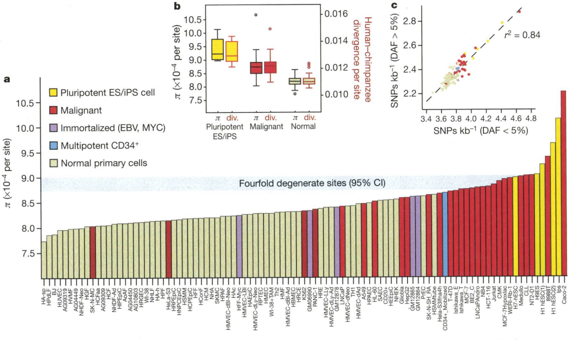


**Figure 7 | Genetic variation in regulatory DNA linked to mutation rate.** **a**, Mean nucleotide diversity ($\pi$, $y$ axis) in DHSs of 97 diverse cell types ($x$ axis) estimated using whole-genome sequencing data from 53 unrelated individuals. Cell types are ordered left-to-right by increasing mean $\pi$. Horizontal blue bar shows 95% confidence intervals on mean $\pi$ in a background model of fourfold degenerate coding sites. Note the enrichment of immortal cells at right. ES, embryonic stem; iPS, induced pluripotent stem. **b**, Mean $\pi$ (left $y$ axis) for pluripotent (yellow) versus malignancy-derived (red) versus normal cells (light green), plotted side-by-side with human–chimpanzee divergence (right $y$ axis) computed on the same groups. Boxes indicate 25–75 percentiles, with medians highlighted. **c**, Both low- and high-frequency derived alleles show the same effect. Density of SNPs in DHSs with derived allele frequency (DAF) <5% ($x$ axis) is tightly correlated ($r^2 = 0.84$) with the same measure computed for higher-frequency derived alleles ($y$ axis). Colour-coding is the same as in panel **a**.

lines exhibit differences in SNP densities but not in allele frequency distribution (Fig. 7c). Collectively, these observations are consistent with increased relative mutation rates in the DHS compartment of immortal cells versus cell types with limited proliferative potential, exposing an unexpected link between chromatin accessibility, proliferative potential and patterns of human variation.

## Discussion

Since their discovery over 30 years ago, DNase I hypersensitive sites have guided the discovery of diverse *cis*-regulatory elements in the human and other genomes. Here we have presented by far the most comprehensive map of human regulatory DNA, revealing novel relationships between chromatin accessibility, transcription, DNA methylation and the occupancy of sequence-specific factors. The wide spectrum of different cell and tissue types covered by our data greatly expands the horizons of cell-selective gene regulation analysis, enabling the recognition of systematic long-distance regulatory patterns, and previously undescribed phenomena such as stereotyping of DHS activation and mutation rate variation in normal versus immortal cells. The extensive resources we have provided should greatly facilitate future analyses, and stimulate new areas of investigation into the organization and control of the human genome. Co-published ENCODE-related papers can be explored online via the *Nature* ENCODE explorer (http://www.nature.com/ENCODE), a specially designed visualization tool that allows users to access the linked papers and investigate topics that are discussed in multiple papers via thematically organized threads.

## METHODS SUMMARY

DNase I hypersensitivity mapping was performed using protocols developed by Duke University[7] or University of Washington[8] on a total of 125 cell types (Supplementary Table 1). Data sets were sequenced to an average depth of 30 million uniquely mapping sequence tags (27–36 bp for University of Washington and 20 bp for Duke University) per replicate. For uniformity of analysis, some cell-type data sets that exceeded 40M tag depth were randomly subsampled to a depth of 30 million tags. Sequence reads were mapped using the Bowtie aligner, allowing a maximum of two mismatches. Only reads mapping uniquely to the genome were used in our analyses. Mappings were to male or female versions of hg19/GRCh37, depending on cell type, with random regions omitted. Data were analysed jointly using a single algorithm[7] (Supplementary Methods) to localize DNase I hypersensitive sites. H3K4me3 ChIP-seq was performed using antibody 9751 (Cell Signaling) on 1% formaldehyde crosslinked samples sheared by Diagenode Bioruptor. Gene expression measurements for each cell type were performed on Affymetrix human exon microarrays. 5C experiments were performed as described[30,31]. Transcription factor recognition motif occurrences within DHSs were defined with FIMO[36] at significance $P < 10^{-5}$ using motif models from the TRANSFAC database.

**Received 15 December 2011; accepted 15 May 2012.**

1. Felsenfeld, G., Boyes, J., Chung, J., Clark, D. & Studitsky, V. Chromatin structure and gene expression. *Proc. Natl Acad. Sci. USA* **93,** 9384–9388 (1996).
2. Gross, D. S. & Garrard, W. T. Nuclease hypersensitive sites in chromatin. *Annu. Rev. Biochem.* **57,** 159–197 (1988).
3. Gaszner, M. & Felsenfeld, G. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nature Rev. Genet.* **7,** 703–713 (2006).
4. Li, Q., Harju, S. & Peterson, K. R. Locus control regions: coming of age at a decade plus. *Trends Genet.* **15,** 403–408 (1999).
5. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genet.* **39,** 311–318 (2007).
6. Hesselberth, J. R. *et al.* Global mapping of protein-DNA interactions *in vivo* by digital genomic footprinting. *Nature Methods* **6,** 283–289 (2009).
7. Boyle, A. P. *et al.* High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132,** 311–322 (2008).
8. John, S. *et al.* Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nature Genet.* **43,** 264–268 (2011).
9. Song, L. *et al.* Open chromatin defined by DNase I and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res.* **21,** 1757–1767 (2010).
10. Harrow, J. *et al.* GENCODE: The reference human genome annotation for the ENCODE project. *Genome Res.* (in the press).
11. Griffiths-Jones, S., Saini, H. K., van Dongen, S. & Enright, A. J. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* **36,** D154–D158 (2008).
12. Farazi, T. A., Spitzer, J. I., Morozov, P. & Tuschl, T. miRNAs in human cancer. *J. Pathol.* **223,** 102–115 (2011).

13. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* http://dx.doi.org/10.1038/nature11233 (this issue).
14. ENCODE Project Consortium.. An integrated encyclopedia of DNA elements in the human genome. *Nature* http://dx.doi.org/10.1038/nature11247 (this issue).
15. Biggin, M. D. Animal transcription networks as highly connected, quantitative continua. *Dev. Cell* **21,** 611–626 (2011).
16. Reddy, P. M., Stamatoyannopoulos, G., Papayannopoulou, T. & Shen, C. K. Genomic footprinting and sequencing of human β-globin locus. Tissue specificity and cell line artifact. *J. Biol. Chem.* **269,** 8287–8295 (1994).
17. Forsberg, E. C., Downs, K. M. & Bresnick, E. H. Direct interaction of NF-E2 with hypersensitive site 2 of the β-globin locus control region in living cells. *Blood* **96,** 334–339 (2000).
18. Talbot, D. & Grosveld, F. The 5′HS2 of the globin locus control region enhances transcription through the interaction of a multimeric complex binding at two functionally distinct NF-E2 binding sites. *EMBO J.* **10,** 1391–1398 (1991).
19. Weisbrod, S. & Weintraub, H. Isolation of a subclass of nuclear proteins responsible for conferring a DNase I-sensitive structure on globin chromatin. *Proc. Natl Acad. Sci. USA* **76,** 630–634 (1979).
20. Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G. & Rauscher, F. J. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev.* **16,** 919–932 (2002).
21. Frietze, S., O'Geen, H., Blahnik, K. R., Jin, V. X. & Farnham, P. J. ZNF274 recruits the histone methyltransferase SETDB1 to the 3′ ends of ZNF genes. *PLoS ONE* **5,** e15082 (2010).
22. Stergachis, A. B., Maclean, B., Lee, K., Stamatoyannopoulos, J. A. & MacCoss, M. J. Rapid empirical discovery of optimal peptides for targeted proteomics. *Nature Methods* **8,** 1041–1043 (2011).
23. Henikoff, S., Henikoff, J. G., Sakai, A., Loeb, G. B. & Ahmad, K. Genome-wide profiling of salt fractions maps physical properties of chromatin. *Genome Res.* **19,** 460–469 (2009).
24. Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148,** 84–98 (2012).
25. Siegfried, Z. *et al.* DNA methylation represses transcription *in vivo. Nature Genet.* **22,** 203–206 (1999).
26. O'Geen, H. *et al.* Genome-wide analysis of KAP1 binding suggests autoregulation of KRAB-ZNFs. *PLoS Genet.* **3,** e89 (2007).
27. Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480,** 490–495 (2011).
28. Rasheed, Z. A., Saleem, A., Ravee, Y., Pandolfi, P. P. & Rubin, E. H. The topoisomerase I-binding RING protein, topors, is associated with promyelocytic leukemia nuclear bodies. *Exp. Cell Res.* **277,** 152–160 (2002).
29. Dahle, Ø., Bakke, O. & Gabrielsen, O. S. c-Myb associates with PML in nuclear bodies in hematopoietic cells. *Exp. Cell Res.* **297,** 118–126 (2004).
30. Dostie, J. *et al.* Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* **16,** 1299–1309 (2006).
31. Sanyal, A., Lajoie, B., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* http://dx.doi.org/10.1038/nature11279 (this issue).
32. Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S. H. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132,** 1049–1061 (2008).
33. Tuan, D., Kong, S. & Hu, K. Transcription of the hypersensitive site HS2 enhancer in erythroid cells. *Proc. Natl Acad. Sci. USA* **89,** 11219–11223 (1992).
34. Neph, S. *et al.* An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* http://dx.doi.org/10.1038/nature11212 (this issue).
35. Vernot, B. *et al.* Personal and population genomics of human regulatory variation. *Genome Res.* (in the press).
36. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27,** 1017–1018 (2011).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank our ENCODE colleagues for many insights into the data types generated by different centres and for help with coordinated analyses. We thank I. Stanaway for assistance with the variation analysis, and many colleagues, particularly F. Urnov, for helpful critiques of the manuscript and figures. This work was funded by National Institutes of Health grants HG004592 (J.A.S.), HG004563 (G.E.C.), GM076036 (J.M.A.) and R01MH084676 (S.R.S), and J.V. is supported by the National Science Foundation Graduate Research Fellowship under grant no. DGE-0718124. N.C.S. is supported by a National Science Foundation Graduate Research Fellowship and the Research Council of Norway. M.T. and K.G. acknowledge funding support from the caBIG In Silico Center of Excellence, NCI/NIH contract no. HHSN261200800001E.

**Author Contributions** Generation of DNase I data was supervised by J.A.S. and G.E.C., with data collection carried out by M.O.D., P.J.S., R.K., D.B., T.K.C., R.S.H., M.D., D.D., E.G., T.K., K.L., F.N., V.R., A. Shafer, S.V., M.W., B.-K.L., D. London, L.S., Zhancheng Z. and Zhuzhu Z. 5C experiments were supervised by J.D. and performed by A. Sanyal. Primary DNase I data processing was performed by R.S., T.S.F., A.K.J. and A.P.R. Hypersensitivity Southern blots and enhancer cloning and transfection experiments were performed by E.M.J., A.K.E., T.F., E.D.N., L.B., D. Lotakis, M.E.S. and Y.Y. and supervised by P.A.N. and G.S. H3K4me3 ChIP-seq experiments were performed by H.W. Primary analysis of DNase I data was performed by R.E.T., R.S. and R.H. Joint analysis of DNase I and transcription factor ChIP-seq data was performed by J.V., S.N., A.B.S. and H.Q. Promoter prediction analysis was performed by R.E.T. DNase I versus DNA methylation analysis was performed by M.T.M. DHS–promoter connectivity analysis was performed by E.R. Integration of DNase I and 5C data was performed by R.H. with assistance from B. Lajoie. DHS stereotyping pattern analysis was performed by E.H. Self-organizing map analysis was performed by N.C.S. and B. Lenhard. MicroRNA analysis was performed by K.G., J.M.S. and M.T. Variation analysis was performed by B.V. and E.R. under direction of S.R.S., J.M.A. and J.A.S. Data interpretation and figure design were performed by J.A.S., R.E.T., J.D.L., V.R.I., G.E.C. and T.S.F. J.A.S., R.E.T., E.R., R.H., J.V., M.T.M., A.B.S., S.J. and N.S. wrote the paper.

**Author Information** DNase I-seq data are available through the UCSC browser, and through the NCBI Gene Expression Omnibus (GEO) data repository under accessions GSE29692 and GSE32970. H3K4me3 data are available through the UCSC browser, and through the NCBI GEO data repository under accession GSE35583. Data for 5C are available through the UCSC browser under accession wgEncodeEH002102. Gene expression data are available through the UCSC browser, and through the NCBI GEO data repository under accessions GSE19090, GSE15805 and GSE17778. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and the online version of the paper is freely available to all readers. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.A.S. (jstam@uw.edu).

doi:10.1038/nature11212

# An expansive human regulatory lexicon encoded in transcription factor footprints

Shane Neph[1]*, Jeff Vierstra[1]*, Andrew B. Stergachis[1]*, Alex P. Reynolds[1]*, Eric Haugen[1], Benjamin Vernot[1], Robert E. Thurman[1], Sam John[1], Richard Sandstrom[1], Audra K. Johnson[1], Matthew T. Maurano[1], Richard Humbert[1], Eric Rynes[1], Hao Wang[1], Shinny Vong[1], Kristen Lee[1], Daniel Bates[1], Morgan Diegel[1], Vaughn Roach[1], Douglas Dunn[1], Jun Neri[1], Anthony Schafer[1], R. Scott Hansen[1,2], Tanya Kutyavin[1], Erika Giste[1], Molly Weaver[1], Theresa Canfield[1], Peter Sabo[1], Miaohua Zhang[3], Gayathri Balasundaram[3], Rachel Byron[3], Michael J. MacCoss[1], Joshua M. Akey[1], M. A. Bender[3,4], Mark Groudine[3,5], Rajinder Kaul[1,2] & John A. Stamatoyannopoulos[1,6]

**Regulatory factor binding to genomic DNA protects the underlying sequence from cleavage by DNase I, leaving nucleotide-resolution 'footprints'. Using genomic DNase I footprinting across 41 diverse cell and tissue types, we detected 45 million transcription factor occupancy events within regulatory regions, representing differential binding to 8.4 million distinct short sequence elements. Here we show that this small genomic sequence compartment, roughly twice the size of the exome, encodes an expansive repertoire of conserved recognition sequences for DNA-binding proteins that nearly doubles the size of the human *cis*-regulatory lexicon. We find that genetic variants affecting allelic chromatin states are concentrated in footprints, and that these elements are preferentially sheltered from DNA methylation. High-resolution DNase I cleavage patterns mirror nucleotide-level evolutionary conservation and track the crystallographic topography of protein–DNA interfaces, indicating that transcription factor structure has been evolutionarily imprinted on the human genome sequence. We identify a stereotyped 50-base-pair footprint that precisely defines the site of transcript origination within thousands of human promoters. Finally, we describe a large collection of novel regulatory factor recognition motifs that are highly conserved in both sequence and function, and exhibit cell-selective occupancy patterns that closely parallel major regulators of development, differentiation and pluripotency.**

ENCODE
Encyclopedia of DNA Elements
nature.com/encode

Sequence-specific transcription factors interpret the signals encoded within regulatory DNA. The discovery of DNase I footprinting over 30 years ago[1] revolutionized the analysis of *cis*-regulatory sequences in diverse organisms, and directly enabled the discovery of the first human sequence-specific transcription factors[2]. Binding of transcription factors to regulatory DNA regions in place of canonical nucleosomes triggers chromatin remodelling, resulting in nuclease hypersensitivity[3]. Within DNase I hypersensitive sites (DHSs), DNase I cleavage is not uniform; rather, punctuated binding by sequence-specific regulatory factors occludes bound DNA from cleavage, leaving footprints that demarcate transcription factor occupancy at nucleotide resolution[1,4] (Fig. 1a). DNase I footprinting has been applied widely to study the dynamics of transcription factor occupancy and cooperativity within regulatory DNA regions of individual genes[5], and to identify cell- and lineage-selective transcriptional regulators[6].

## Regulatory DNA is populated with DNase I footprints

To map DNase I footprints comprehensively within regulatory DNA, we adapted digital genomic footprinting[4] to human cells. The ability to resolve DNase I footprints sensitively and precisely is critically dependent on the local density of mapped DNase I cleavages (Supplementary Fig. 1a–d), and efficient footprinting of a large genome such as human requires substantial concentration of DNase I cleavages within the small fraction (~1–3%) of the genome contained in DNase I-hypersensitive regions. We selected highly enriched DNase I cleavage libraries from 41 diverse cell types in which 53–81% of DNase I cleavage sites localized to DNase I-hypersensitive regions[7] (Supplementary Table 1), representing nearly tenfold higher signal-to-noise ratio than previous results from yeast[4], and two- to fivefold greater enrichment than achieved using end-capture of single DNase I cleavages[8,9]. We then performed deep sequencing of these libraries, and obtained 14.9 billion Illumina sequence reads, 11.2 billion of which mapped to unique locations in the human genome (Supplementary Table 1). We achieved an average sequencing depth of ~273 million DNase I cleavages per cell type that enabled extensive and accurate discrimination of DNase I footprints.

To detect DNase I footprints systematically, we implemented a detection algorithm based on the original description of quantitative DNase I footprinting[1] (Supplementary Methods). We identified an average of ~1.1 million high-confidence (false discovery rate (FDR) of 1%) footprints per cell type (range 434,000 to 2.3 million; Supplementary Table 1), and collectively 45,096,726 6–40-base pair (bp) footprint events across all cell types. We resolved cell-selective footprint patterns to reveal 8.4 million distinct elements with a footprint, each occupied in one or more cell type. At least one footprint was found in >75% of DHSs (Supplementary Fig. 1c, d and Supplementary Table 2), with detection strongly dependent on the number of mapped DNase I cleavages within each DHS. 99.8% of DHSs with >250 mapped DNase I cleavages contained at least one footprint, indicating that DHSs are not simply open or nucleosome-free chromatin features, but are constitutively populated with DNase I footprints. Modelling

---

[1]Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. [2]Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, Washington 98195, USA. [3]Fred Hutchison Cancer Research Center, Seattle, Washington 98109, USA. [4]Department of Pediatrics, University of Washington, Seattle, Washington 98195, USA. [5]Department of Radiation Oncology, Department of Medicine, University of Washington, Seattle, Washington 98195, USA. [6]Division of Oncology, Department of Medicine, University of Washington, Seattle, Washington 98195, USA.
*These authors contributed equally to this work.

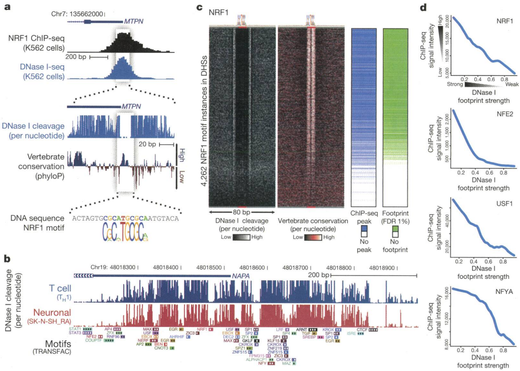


**Figure 1 | Parallel profiling of genomic regulatory factor occupancy across 41 cell types.** **a**, DNase I footprinting of K562 cells identifies the individual nucleotides within the *MTPN* promoter that are bound by NRF1. **b**, Example locus harbouring eight clearly defined DNase I footprints in T-helper type 1 ($T_H1$) and SK-N-SH_RA cells, with TRANSFAC database motif instances indicated below. **c**, Heat maps showing per-nucleotide DNase I cleavage (left) and vertebrate conservation by phyloP (right) for 4,262 NRF1 motifs within K562 DHSs ranked by the local density of DNase I cleavages. Green ticks indicate the presence of DNase I footprints over motif instances. Blue ticks indicate the presence of ChIP-seq peaks over the motif instances. **d**, Lowess regression of NRF1, USF1, NFE2 and NFYA K562 ChIP-seq signal intensities versus DNase I footprinting occupancy (footprint occupancy score) at K562 DNase I footprints containing NRF1, USF, NFE2 and NFYA motifs.

DNase I cleavage patterns using empirically derived intrinsic DNA cleavage propensities for DNase I showed that only a miniscule fraction (0.24%) of discovered 1% FDR footprints from cell and tissue samples could be caused by inherent DNase I sequence specificity (Supplementary Methods).

DNase I footprints were distributed throughout the genome, including intergenic regions (45.7%), introns (37.7%), upstream of transcriptional start sites (TSSs, 8.9%), and in 5′ and 3′ untranslated regions (UTRs, 1.4% and 1.3%, respectively; Supplementary Fig. 2a, b). DNase I footprints were enriched in promoters (3.6-fold; $P < 2.2 \times 10^{-16}$; Binomial test) and 5′ UTRs (2.4-fold; $P < 2.2 \times 10^{-16}$; Binomial test), commensurate with high DNase I cleavage densities observed in these regions. We found that 2.0% of footprints localized within exons, raising the possibility that occupancy by DNA binding proteins could further restrict sequence diversity within coding DNA, thus superimposing an unexpected layer of constraint on codon usage.

## Footprints are quantitative markers of factor occupancy

We next examined the correspondence between DNase I footprints and known regulatory factor recognition sequences within DNase I hypersensitive chromatin. Comprehensive scans of DNase I hypersensitive regions for high-confidence matches to all recognized transcription factor motifs in the TRANSFAC[10] and JASPAR[11] databases revealed a striking enrichment of motifs within footprints ($P \approx 0$, $z$-score = 204.22 for TRANSFAC; $z$-score = 169.88 for JASPAR; Fig. 1b and Supplementary Fig. 3).

To quantify the occupancy at transcription factor recognition sequences within DHSs genome-wide, we computed for each instance a footprint occupancy score (FOS) relating the density of DNase I cleavages within the core recognition motif to cleavages in the immediately flanking regions (Supplementary Methods). The FOS can be used to rank motif instances by the 'depth' of the footprint at that position, and is expected to provide a quantitative measure of factor occupancy[1]. To examine this relationship for a well-studied sequence-specific regulator (NRF1; ref. 12), we plotted DNase I cleavage patterns surrounding all 4,262 NRF1 motifs contained within DHSs and ranked these by FOS. Whereas only a subset of these motif instances (2,351) coincided with high-confidence footprints, the vast majority of NRF1 motif instances in DNase I footprints (89%) overlapped reproducible sites of NRF1 occupancy identified by chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) (Fig. 1c). In parallel, we analysed nucleotide-level evolutionary conservation patterns around NRF1-binding sites, revealing that FOS closely parallels phylogenetic conservation within the core motif region, indicating strong selection on factor occupancy (Fig. 1c). We observed a nearly monotonic relationship between FOS and ChIP-seq signal intensities at NRF1-binding sites within DNase I footprints of K562 cells (Fig. 1d). Similarly strong correlations between footprint occupancy and either ChIP-seq signal or phylogenetic conservation were evident for diverse factors (Fig. 1d and Supplementary Fig. 4a–d). We found that footprint occupancy and nucleotide-level conservation correlated for 80% of all transcription factor motifs in the TRANSFAC database, of which 50% were statistically significant ($P < 0.05$; Supplementary Methods). This relationship between footprint occupancy and conservation is most readily explained by evolutionary selection on factor occupancy, with higher conservation of higher affinity binding sites. Taken together, these results indicate that footprint occupancy provides a quantitative measure of sequence-specific regulatory factor occupancy that closely parallels evolutionary constraint and ChIP-seq signal intensity.

To validate the potential for selective binding of footprints by factors predicted on the basis of motif-to-footprint matching, we developed an approach to quantify specific occupancy in the context of a complex transcription factor milieu using targeted mass spectrometry (DNA interacting protein precipitation or DIPP; Methods). Using DIPP, we affirmed specific binding by several different classes of transcription factor (Supplementary Fig. 5a–e). Together with the analysis of ChIP-seq data described above, these results indicate that the localization of transcription factor recognition motifs within DNase I footprints can accurately illuminate the genomic protein occupancy landscape.

## Footprints harbour functional SNVs and lack methylation

The potential for single nucleotide variants (SNVs) within a transcription factor recognition sequence to abrogate binding of its cognate factor is well known[13]. The depth of sequencing performed in the context of our footprinting experiments provided hundreds- to thousands-fold coverage of most DHSs, enabling precise quantification of allelic imbalance within DHSs harbouring heterozygous variants. We scanned all DHSs for heterozygous SNVs identified by the 1000 Genomes Project[14] and measured, for each DHS containing a single heterozygous variant, the proportion of reads from each allele. We identified likely functional variants conferring significant allelic imbalance in chromatin accessibility and analysed their distribution relative to DNase I footprints. This analysis revealed significant enrichment ($P < 2.2 \times 10^{-16}$; Fisher's exact test) of such variants within DNase I footprints (Supplementary Fig. 6). For example, rs4144593 is a common T-to-C (T/C) variant that lies within a DHS on chromosome 9. This variant falls on a high-information position within a footprint containing an NF1/CTF1 motif and substantially disrupts footprinting of this motif, resulting in allelic imbalance in chromatin accessibility (Fig. 2a).

Protein–DNA interactions are also sensitive to cytosine methylation[15,16]. Comparing DNase I footprints and whole-genome bisulphite sequencing methylation data from pulmonary fibroblasts (IMR90), we found that CpG dinucleotides contained within DNase I footprints were significantly less methylated than CpGs in non-footprinted regions of the same DHS (Mann–Whitney $U$-test; $P < 2.2 \times 10^{-16}$; Fig. 2b). Footprints therefore seem to be selectively sheltered from DNA methylation, indicating a widespread connection between regulatory factor occupancy and nucleotide-level patterning of epigenetic modifications.

## Transcription factor structure is imprinted on the genome

We observed surprisingly heterogeneous base-to-base variation in DNase I cleavage rates within the footprinted recognition sequences of different regulatory factors. And yet, the per site cleavage profiles for individual factors were highly stereotyped, with nearly identical local cleavage patterns at thousands of genomic locations (Supplementary Fig. 7). This raised the possibility that DNase I cleavage patterns may provide information concerning the morphology of the DNA–protein interface. We obtained the available DNA–protein co-crystal structures for human transcription factors, and mapped aggregate DNase I cleavage patterns at individual nucleotide positions onto the DNA backbone of the co-crystal model. Figure 3a and Supplementary Fig. 8a show two examples: USF1 (ref. 17) and SRF[18]. For both factors, DNase I cleavage patterns clearly parallel the topology of the protein–DNA interface, including a marked depression in DNase I cleavage at nucleotides involved in protein–DNA contact, and increased cleavage at exposed nucleotides such as those within the central pocket of the leucine zipper. These data show that nucleotide-level aggregate DNase I cleavage patterns reflect fundamental features of the protein–DNA interaction interface at unprecedented resolution.

We next asked how these patterns related to evolutionary conservation. Plotting nucleotide-level aggregate DNase I cleavage in parallel with per-nucleotide vertebrate conservation calculated by phyloP[19] revealed striking antiparallel patterning of cleavage versus conservation across nearly all motifs examined (six representative examples are shown in Fig. 3b and Supplementary Fig. 8b). Notably, conservation is not limited to only DNA contacting protein residues, but exhibits graded changes that mirror DNase I accessibility across the entirety of the protein–DNA interface (Supplementary Figs 8c, d). Taken together, these results imply that regulatory DNA sequences have evolved to fit the continuous morphology of the transcription factor–DNA binding interface.

## A ~50-bp footprint localizes transcription initiation

Transcription initiation requires the binding of multi-protein complexes that position RNA polymerase II[20–23]. Using a modified footprint detection algorithm designed to detect larger features (Supplementary Methods), we scanned the regions upstream from GENCODE TSSs and identified highly stereotyped ~80-bp chromatin structure comprising a prominent ~50-bp central DNase I footprint, flanked symmetrically by ~15-bp regions of uniformly elevated levels of DNase I cleavage (Fig. 4a). Alignment of per-nucleotide DNase I cleavage profiles from 5,041 prominent footprints mapped in different K562 promoters highlights the homogeneous, nearly invariant nature of the structure (Fig. 4b).

Plotting evolutionary conservation in parallel with DNase I cleavage revealed two distinct peaks in evolutionary conservation within the central footprint (Fig. 4c) compatible with binding sites for paired

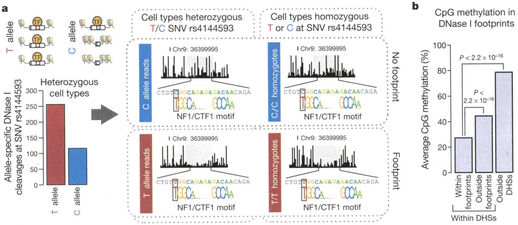


**Figure 2 | DNase I footprints mark sites of *in vivo* protein occupancy.** **a**, Schematic and plots showing the effect of T/C SNV rs4144593 on protein occupancy and chromatin accessibility. The $y$ axis of the bar graph shows the number of DNase I cleavage events containing either the T or C allele. Middle plots show T or C allele-specific DNase I cleavage profiles from ten cell lines heterozygous for the T/C alleles at rs4144593. Right plots show DNase I cleavage profiles from 18 cell lines homozygous for the C allele at rs4144593 and one cell line homozygous for the T allele at rs4144593. Cleavage plots are cut off at 60% cleavage height. **b**, The average CpG methylation within IMR90 DNase I footprints, IMR90 DHSs (but not in footprints) and non-hypersensitive genomic regions in IMR90 cells. CpG methylation is significantly depleted in DNase I footprints ($P < 2.2 \times 10^{-16}$, Mann–Whitney $U$-test).

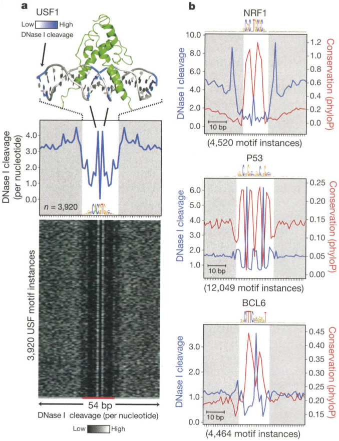


**Figure 3 | Footprint structure parallels transcription factor structure and is imprinted on the human genome.** **a**, The co-crystal structure of upstream stimulatory factor (USF1) bound to its DNA ligand is juxtaposed above the average nucleotide-level DNase I cleavage pattern (blue) at motif instances of USF in DNase I footprints. Nucleotides that are sensitive to cleavage by DNase I are coloured blue on the co-crystal structure. The motif logo generated from USF DNase I footprints is displayed below the DNase I cleavage pattern. Below is a randomly ordered heat map showing the per-nucleotide DNase I cleavage for each motif instance of USF in DNase I footprints. **b**, The per-base DNase I hypersensitivity (blue) and vertebrate phylogenetic conservation (red) for all DNase I footprints in dermal fibroblasts matching three well-annotated transcription factor motifs. The white box indicates width of consensus motif. The number of motif occurrences within DNase I footprints is indicated below each graph.

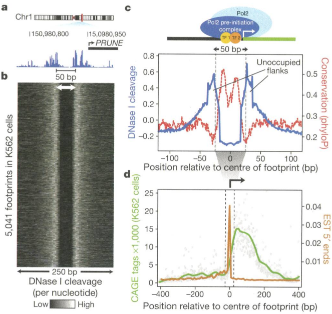


**Figure 4 | A highly stereotyped chromatin structural motif marks sites of transcription initiation in human promoters.** **a**, A 35–55-bp footprint is the predominant feature of many promoter DHSs and is in tight spatial coordination with the transcription start site. **b**, Heat map of the per-nucleotide DNase I cleavage pattern at 5,041 instances of this stereotypical footprint in K562 cells. **c**, Aggregate per-base DNase I cleavage profile (blue line) and mean per-nucleotide conservation score (phyloP) surrounding instances of this stereotypical footprint in K562 cells (red dashed line). **d**, Aggregate strand corrected CAGE sequencing data (green line) and the average nearest 5′ end of a spliced EST (orange line) surrounding instances of this stereotypical footprint in K562 cells.

canonical sequence-specific transcription factors. The density of capped analysis of gene expression (CAGE) tags (Fig. 4d; green line) and 5′ ends of expressed sequenced tags (ESTs) (Fig. 4d; orange line) relative to the central ~50-bp footprint revealed that, at the vast majority of promoters, RNA transcript initiation localized precisely within the stereotyped footprint. It is notable that the location of this footprint is often offset, typically 5′, from many GENCODE-annotated TSSs. This probably derives from the incomplete nature of many of the 5′ transcript ends used to define TSSs[24].

These data together define a new high-resolution chromatin structural signature of transcription initiation and the interaction of the pre-initiation complex with the core promoter. Indeed, chromatin occupancy of TATA-binding protein (TBP), a critical component of the pre-initiation complex, is maximal precisely over the centre of the 50-bp footprint region (Supplementary Fig. 9a). Sequence analysis of the two conservation peaks within the 50-bp footprint identified motifs for GC-box-binding proteins such as SP1 and, less frequently, other general transcription factors (though with the notable absence of TATA motifs) (Supplementary Fig. 9b), indicating that TBP (and potentially other pre-initiation complex components) interacts preferentially with general transcriptional factors bound to GC-box-like features in the central footprinted region. The results are therefore consistent with a model in which a limited number of sequence-specific factors function both to prime the chromatin template for recruitment of RNA polymerase II and to guide transcriptional positioning.

## Distinguishing indirect transcription factor occupancy

Many transcriptional regulators are posited to interact indirectly with the DNA sequence of some target sites through mechanisms such as tethering[25]. Approaches such as ChIP-seq detect chromatin occupancy, but cannot by themselves distinguish sites of direct DNA binding from non-canonical indirect binding. We therefore asked whether DNase I footprint data could illuminate ChIP-seq-derived occupancy profiles by differentiating directly bound factors from indirect binding events. We first partitioned ChIP-seq peaks from each of 38 ENCODE transcription factors[26] mapped in K562 cells into three categories of predicted sites: ChIP-seq peaks containing a compatible footprinted motif (directly bound sites); ChIP-seq peaks lacking a compatible motif or footprint (indirectly bound sites); and ChIP-seq peaks overlying a compatible motif lacking a footprint (indeterminate sites). Predicted indirect sites showed significantly reduced ChIP-seq signal compared with predicted directly bound sites (Supplementary Fig. 10), consistent with lack of direct crosslinking to DNA (and therefore reduced ChIP efficiency). Indeterminate sites exhibited low ChIP-seq signal and were therefore excluded from further analysis (Supplementary Fig. 10).

The fraction of ChIP-seq peaks predicted to represent direct versus indirect binding varied widely between different factors, ranging from nearly complete direct sequence-specific binding (for example, CTCF), to nearly complete indirect binding (for example, TBP; Supplementary Fig. 11). In many cases factors that preferentially engage in direct DNA binding at distal sites show predominantly indirect occupancy in promoter regions and vice versa (Supplementary Fig. 12a, b).

Next, we analysed the frequency with which indirectly bound sites of one transcription factor coincided with directly bound sites of a second factor, indicative of protein–protein interactions (for example, tethering). This analysis recovered many known protein–protein interactions, such as CTCF–YY1 and TAL1–GATA1 (ref. 27), as well as many novel associations (Fig. 5). We observed enrichment for NFE2 indirect interactions at promoter-bound USF2 sites, compatible with their known interaction[28]. At distal sites, we observed the opposite, with NFE2 predominantly directly bound accompanied by USF2 indirect peaks (Supplementary Fig. 12a, b), indicating the possibility of a reciprocal or looping mechanism. Notably, directly bound promoter-predominant transcription factors were enriched for co-localization with indirect peaks compared to distal regions (Supplementary Fig. 13a, b). These results suggest that combining DNase I footprinting with ChIP-seq has the potential to expose a previously unappreciated landscape of complex transcription factor occupancy modes.

## Footprints encode an expansive *cis*-regulatory lexicon

Since the discovery of the first sequence-specific transcription factor[29], considerable effort has been devoted to identifying the cognate recognition sequences of DNA-binding proteins[30,31]. Despite these efforts, high-quality motifs are available for only a minority of the >1,400 human transcription factors with predicted sequence-specific DNA binding domains[32].

We reasoned that the genomic sequence compartment defined by DNase I footprints in a given cell type ideally should contain much, if not all, of the factor recognition sequence information relevant for that cell type. Consequently, applying *de novo* motif discovery to the footprint compartments gleaned from multiple cell types should greatly expand our current knowledge of biologically active transcription factor binding motifs.

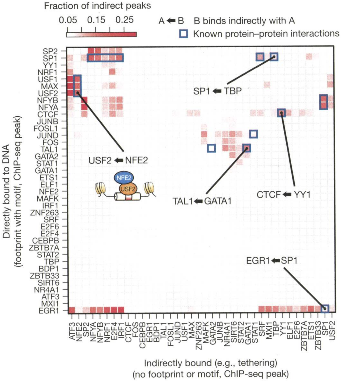


**Figure 5 | Distinguishing direct and indirect binding of transcription factors.** Heat map of the enrichment of pairs of transcription factors in a direct–indirect association. Direct peaks are defined by ChIP occupancy accompanied by a footprint overlapping a compatible motif. Indirect peaks do not have a compatible motif. The colour of each cell is determined by the fraction of indirect peaks that co-localize with the direct peaks of another factor.

We performed unbiased *de novo* motif discovery within the footprints identified in each of the 41 cell types that yielded 683 unique motif models (Fig. 6a and Supplementary Methods). We compared these models with the universe of experimentally grounded motif models in the TRANSFAC, JASPAR and UniPROBE[33] databases. Owing to the redundancy of motif models contained within these databases, we first collapsed all duplicate models (Supplementary Methods). A total of 394 of the 683 (58%) *de novo* motifs matched distinct experimentally grounded motif models, accounting collectively for 90% of all unique entries across the three databases (Fig. 6b and Supplementary Fig. 14a–c). The wholesale *de novo* derivation of the vast majority of known regulatory factor recognition sequences from the small genomic compartment defined by DNase I footprints highlights the marked concentration of regulatory information encoded within this sequence space.

Notably, 289 of the footprint-derived motifs were absent from major databases (Fig. 6b and Supplementary Fig. 14d). These novel motifs populate millions of DNase I footprints (Fig. 6c), and show features of *in vivo* occupancy and evolutionary constraint similar to motifs for known regulators, including marked anti-correlation with nucleotide-level vertebrate conservation (Figs 3b, 6e and Supplementary Figs 8 and 15a).

To test whether novel motifs were functionally conserved in an evolutionarily distant mammal, we analysed DNase I cleavage patterns around human novel motifs mapped within DHSs assayed in primary mouse liver tissue (Fig. 6e, f and Supplementary Fig. 15a, b). This analysis demonstrated that many novel motifs show nearly identical DNase I footprint patterns in both human cells and mouse liver, indicating that these novel motifs correspond to evolutionarily conserved transcriptional regulators that are functional in both mouse and human.

Given the conservation of protein occupancy in a distant mammal, we assessed whether the novel motifs are under selection in human populations by analysing nucleotide diversity across all motif instances found within accessible chromatin. Using high-quality genomic sequence data from 53 unrelated individuals[34] (Supplementary Table 4), we calculated the average nucleotide diversity[35] for each individual motif space (Supplementary Fig. 15c). Reduced diversity levels are indicative of functional constraint, through the elimination of deleterious alleles from the population by natural selection. We found that novel motifs are collectively under strong purifying selection in human populations. On average, the new motifs are more constrained than most motifs found in the major databases (Fig. 6d and Supplementary Fig. 15c), even after exclusion of motifs containing highly mutable CpG dinucleotides, which underlie the marked increase in nucleotide diversity seen with a subset of known motifs (Supplementary Fig. 15c, right). Collectively, these results demonstrate that DNase I footprints encode an expansive *cis*-regulatory lexicon encompassing both known transcription factor recognition sequences and novel motifs that are functionally conserved in mouse and bear strong signatures of ongoing selection in humans.

## Novel motif occupancy parallels regulators of cell fate

Cell-selective gene regulation is mediated by the differential occupancy of transcriptional regulatory factors at their cognate *cis*-acting elements. For example, the nerve growth factor gene *VGF* is selectively expressed only within neuronal cells (Fig. 7a), presumably due to the repressive action of the transcriptional regulator NRSF (also called REST) at the *VGF* promoter in non-neuronal cell types[36]. Although *VGF* is expressed only in neuronal cells, its promoter is DNase I-hypersensitive in most cell types (not shown). Examination of nucleotide-level cleavage patterns within the *VGF* promoter exposes its fundamental *cis*-regulatory logic, coordinated by the transcriptional

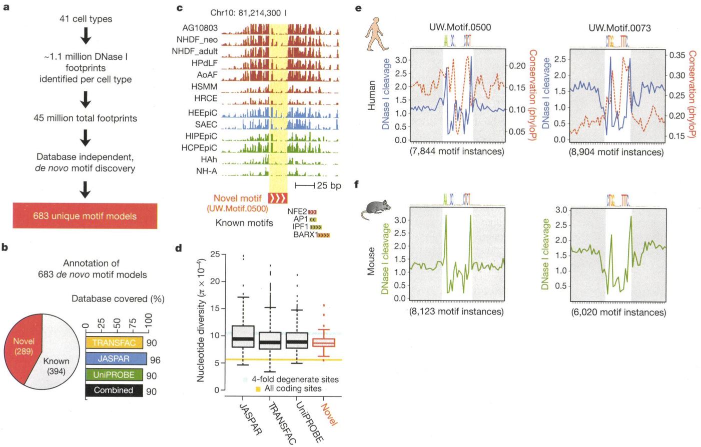


**Figure 6 | *De novo* motif discovery expands the human regulatory lexicon.** **a**, Overview of *de novo* motif discovery using DNase I footprints. **b**, Annotation of the 683 *de novo*-derived motif models using previously identified transcription factor motifs. A total of 394 of these *de novo*-derived motifs match a motif annotated within the TRANSFAC, JASPAR or UniPROBE databases, whereas 289 are novel motifs (pie chart). The *de novo* consensus matching TRANSFAC, JASPAR or UniPROBE sequences cover the majority of each database (bar chart). **c**, Example of a DNase I footprint found in multiple cell types that is annotated solely by one of the novel *de novo*-derived motifs. **d**, Box-and-whisker plot comparing the average nucleotide diversity at instances of the 289 novel *de novo*-derived motif models to instances of motifs present in databases of known specificities (*x* axis). The box defines the 25% and 75% percentiles and the whiskers display 1.5 times the inner quartile range of the distribution of $\pi$ values in each respective database. The blue bar indicates the average nucleotide diversity ($\pi$) at fourfold degenerate coding sites (width is equal to 95% confidence interval); gold bar indicates $\pi$ at all coding sites (width is equal to 95% confidence interval). **e**, Phylogenetic conservation (red dashed) and per-base DNase I hypersensitivity (blue) for all DNase I footprints in dermal fibroblast cells matching two novel *de novo*-derived motifs. The white box indicates width of consensus motif. **f**, Per-nucleotide mouse liver DNase I cleavage patterns at occurrences of the motifs in **e** at DNase I footprints identified in mouse liver.

regulators NRSF, SP1, USF1 and NRF1. Whereas the NRSF motif is tightly occupied in non-neuronal cells, in neuronal cells, NRSF repression is relieved, and recognition sites for the positive regulators USF1 and SP1 become highly occupied, resulting in *VGF* expression. These data collectively illustrate the power of genomic footprinting to resolve differential occupancy of multiple regulatory factors in parallel at nucleotide resolution.

We next extended this paradigm using genome-wide DNase I footprints across 12 functionally distinct cell types to identify both known and novel factors showing highly cell-specific occupancy patterns. To calculate the footprint occupancy of a motif, we enumerated for each motif and cell type the number of motif instances encompassed within DNase I footprints and normalized this by the total number of DNase I footprints in that cell type. Figure 7b shows a heat-map representation of cell-selective occupancy at motifs for 60 known transcriptional regulators and for 29 novel motifs. This approach appropriately identified a number of known cell-selective transcriptional regulators including: (1) the pluripotency factors OCT4 (also called POU5F1), SOX2, KLF4 and NANOG in human embryonic stem cells[37]; (2) the myogenic factors MEF2A and MYF6 in skeletal myocytes[38]; and (3) the erythrogenic regulators GATA1, STAT1 and STAT5A in erythroid cells[39–41] (Fig. 7b).

Many of the footprint-derived novel motifs displayed markedly cell-selective occupancy patterns highly similar with the aforementioned well-established regulators. This suggests that many novel motifs correspond to recognition sequences for important but uncharacterized regulators of fundamental biological processes. Notably, both known and novel motifs with high cell-selective occupancy predominantly localized to distal regulatory regions (Fig. 7c), further highlighting the role of distal regulation in developmental and cell-selective processes[42,43].

## Perspective

We describe an expansive map of regulatory factor occupancy at millions of precisely demarcated sequence elements across the human genome revealed by genomic DNase I footprinting applied to a wide spectrum of cell types. These elements collectively define a highly information-rich genomic sequence compartment that encodes the recognition landscape of hundreds of DNA-binding proteins. This compartment has been extensively shaped by evolutionary forces to match closely the physical properties of its cognate interacting proteins. Mining footprint sequences for recognition motifs has nearly doubled the human *cis*-regulatory lexicon, exposing a previously hidden trove of elements with evolutionary, structural and

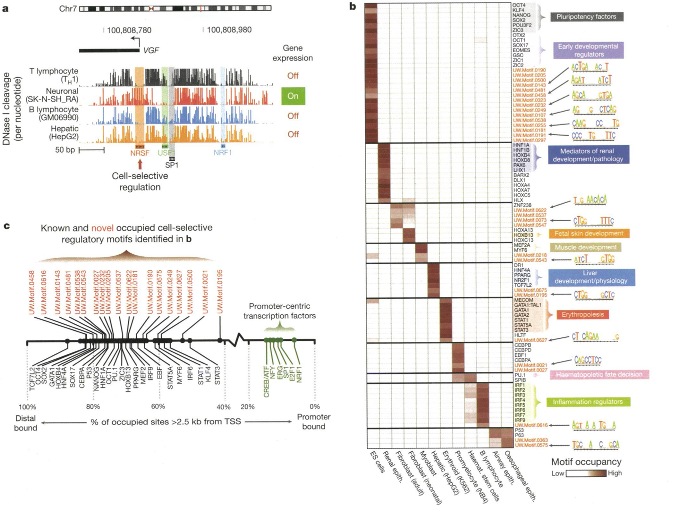


**Figure 7 | Multi-lineage DNase I footprinting reveals cell-selective gene regulators. a,** Comparative footprinting of the nerve growth factor gene (*VGF*) promoter in multiple cell types reveals both conserved (NRF1, USF1 and SP1) and cell-selective (NRSF) DNase I footprints. **b**, Shown is a heat map of footprint occupancy computed across 12 cell types (columns) for 89 motifs (rows), including well-characterized cell/tissue-selective regulators, and novel *de novo*-derived motifs (red text). The motif models for some of these novel *de novo*-derived motifs are indicated next to the heat map. **c**, The proportion of motif instances in DNase I footprints within distal regulatory regions for known (black) and novel (red) cell-type-specific regulators in **b** is indicated. Also noted are these values for a small set of known promoter-proximal regulators (green). ES, embryonic stem.

functional profiles that parallel the collections of experimentally derived genomic regulators brought to light during the past 30 years. Because the ability to resolve footprints is dependent on sequencing depth, and the sequencing level of DNase I cleavage events in most DHSs is not saturating (even in cell types with >500 million mapped unique DNase I cleavages), the present study, although extensive in many respects, represents only an initial foray into this biologically rich space. Identification of the cognate DNA-binding proteins for novel recognition sequences presents a significant challenge, although one that can be addressed with confidence using emerging technologies and our extensive experimental data demonstrating both occupancy *in vivo* and strong evolutionary signatures of function. On a broader level, the approach that we describe here can, in principle, be applied to derive the *cis*-regulatory lexicon of any organism. We anticipate that the extensive new resources we describe, particularly in combination with other ENCODE data, will help to advance many aspects of human gene regulation research. Co-published ENCODE-related papers can be explored online via the *Nature* ENCODE explorer (http://www.nature.com/ENCODE), a specially designed visualization tool that allows users to access the linked papers and investigate topics that are discussed in multiple papers via thematically organized threads.

## METHODS SUMMARY

DNase I digestion and high-throughput sequencing were performed on intact human nuclei from various cell types, following published methods[4,44]. Briefly, roughly 10 million cells were grown in appropriate culture media and nuclei were extracted using NP-40 in an isotonic buffer. The NP-40 detergent was removed and the nuclei were incubated for 3 min at 37 °C with limiting concentrations of the DNA endonuclease, DNase I (Sigma) supplemented with $Ca^{2+}$ and $Mg^{2+}$. The digestion was stopped with EDTA and the samples were treated with proteinase K. The small 'double-hit' fragments (<500 bp) were recovered by sucrose ultra-centrifugation, end-repaired and ligated with adapters compatible with the Illumina sequencing platform. High-quality libraries from each cell type were sequenced on the Illumina platform to an average depth of 273 million uniquely mapping single-end tags. The sequencing tags were aligned to the human reference genome and per-nucleotide cleavage counts were generated by summing the 5′ ends of the aligned sequencing tags at each position in the genome. FDR 1% DNase I footprints were identified using an iterative search method based on optimization of the footprint occupancy score. *De novo* motif discovery was performed using a full enumeration algorithm.

**Received 11 December 2011; accepted 10 May 2012.**

1. Galas, D. J. & Schmitz, A. DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res.* **5,** 3157–3170 (1978).
2. Dynan, W. S. & Tjian, R. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell* **35,** 79–87 (1983).

3. Gross, D. S. & Garrard, W. T. Nuclease hypersensitive sites in chromatin. *Annu. Rev. Biochem.* **57,** 159–197 (1988).
4. Hesselberth, J. R. *et al.* Global mapping of protein-DNA interactions *in vivo* by digital genomic footprinting. *Nature Methods* **6,** 283–289 (2009).
5. Thanos, D. & Maniatis, T. Virus induction of human IFNβ gene expression requires the assembly of an enhanceosome. *Cell* **83,** 1091–1100 (1995).
6. Tsai, S. F. *et al.* Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. *Nature* **339,** 446–451 (1989).
7. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* http://dx.doi.org/10.1038/nature11232 (this issue).
8. Sabo, P. J. *et al.* Discovery of functional noncoding elements by digital analysis of chromatin structure. *Proc. Natl Acad. Sci. USA* **101,** 16837–16842 (2004).
9. Boyle, A. P. *et al.* High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132,** 311–322 (2008).
10. Matys, V. *et al.* TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* **34,** D108–D110 (2006).
11. Bryne, J. C. *et al.* JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res.* **36,** D102–D106 (2008).
12. Chan, J. Y., Han, X. L. & Kan, Y. W. Cloning of Nrf1, an NF-E2-related transcription factor, by genetic selection in yeast. *Proc. Natl Acad. Sci. USA* **90,** 11371–11375 (1993).
13. Rockman, M. V. & Wray, G. A. Abundant raw material for cis-regulatory evolution in humans. *Mol. Biol. Evol.* **19,** 1991–2004 (2002).
14. A map of human genome variation from population-scale sequencing. *Nature* **467,** 1061–1073 (2010).
15. Tate, P. H. & Bird, A. P. Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr. Opin. Genet. Dev.* **3,** 226–231 (1993).
16. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462,** 315–322 (2009).
17. Ferré-D'Amaré, A. R., Pognonec, P., Roeder, R. G. & Burley, S. K. Structure and function of the b/HLH/Z domain of USF. *EMBO J.* **13,** 180–189 (1994).
18. Pellegrini, L., Tan, S. & Richmond, T. J. Structure of serum response factor core bound to DNA. *Nature* **376,** 490–498 (1995).
19. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20,** 110–121 (2010).
20. Pugh, B. F. & Tjian, R. Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes Dev.* **5,** 1935–1945 (1991).
21. Kim, T. H. *et al.* A high-resolution map of active promoters in the human genome. *Nature* **436,** 876–880 (2005).
22. Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56,** 549–561 (1989).
23. Kim, T. K. *et al.* Trajectory of DNA in the RNA polymerase II transcription preinitiation complex. *Proc. Natl Acad. Sci. USA* **94,** 12268–12273 (1997).
24. Affymetrix/Cold Spring Harbor Laboratory ENCODE Transcriptome Project. Post-transcriptional processing generates a diversity of 5′-modified long and short RNAs. *Nature* **457,** 1028–1032 (2009).
25. Biddie, S. C. *et al.* Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. *Mol. Cell* **43,** 145–155 (2011).
26. ENCODE Project Consortium.. An integrated encyclopedia of DNA elements in the human genome. *Nature* http://dx.doi.org/10.1038/nature11247 (this issue).
27. Wadman, I. A. *et al.* The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *EMBO J.* **16,** 3145–3157 (1997).
28. Zhou, Z. *et al.* USF and NF-E2 cooperate to regulate the recruitment and activity of RNA polymerase II in the β-globin gene locus. *J. Biol. Chem.* **285,** 15894–15905 (2010).
29. Gilbert, W. & Müller-Hill, B. Isolation of the *lac* repressor. *Proc. Natl Acad. Sci. USA* **56,** 1891–1898 (1966).
30. Badis, G. *et al.* Diversity and complexity in DNA recognition by transcription factors. *Science* **324,** 1720–1723 (2009).
31. Mukherjee, S. *et al.* Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. *Nature Genet.* **36,** 1331–1339 (2004).
32. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nature Rev. Genet.* **10,** 252–263 (2009).
33. Newburger, D. E. & Bulyk, M. L. UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. *Nucleic Acids Res.* **37,** D77–D82 (2009).
34. Drmanac, R. *et al.* Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* **327,** 78–81 (2010).
35. Nei, M. & Li, W. H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl Acad. Sci. USA* **76,** 5269–5273 (1979).
36. Schoenherr, C. J. & Anderson, D. J. The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science* **267,** 1360–1363 (1995).
37. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–872 (2007).
38. Yun, K. & Wold, B. Skeletal muscle determination and differentiation: story of a core regulatory network and its context. *Curr. Opin. Cell Biol.* **8,** 877–889 (1996).
39. Pevny, L. *et al.* Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* **349,** 257–260 (1991).
40. Socolovsky, M. *et al.* Ineffective erythropoiesis in $Stat5a^{-/-}$ $5b^{-/-}$ mice due to decreased survival of early erythroblasts. *Blood* **98,** 3261–3273 (2001).
41. Halupa, A. *et al.* A novel role for STAT1 in regulating murine erythropoiesis: deletion of STAT1 results in overall reduction of erythroid progenitors and alters their distribution. *Blood* **105,** 552–561 (2005).
42. Treisman, R. & Maniatis, T. Simian virus 40 enhancer increases number of RNA polymerase II molecules on linked. *DNA* **315,** 73–75 (1985).
43. Grosveld, F., van Assendelft, G. B., Greaves, D. R. & Kollias, G. Position-independent, high-level expression of the human β-globin gene in transgenic mice. *Cell* **51,** 975–985 (1987).
44. Sabo, P. J. *et al.* Genome-scale mapping of DNase I sensitivity *in vivo* using tiling DNA microarrays. *Nature Methods* **3,** 511–518 (2006).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was supported by National Institutes of Health (NIH) grants HG004592 (J.A.S.) and RC2HG005654 (J.A.S. and M.G.). J.V. is supported by a National Science Foundation Graduate Research Fellowship under grant no. DGE-071824. Additional support was provided in part by the University of Washington Proteomics Resource (UWPR95794). We thank F. Urnov for critical reading of the manuscript and many discussions, and S. Thomas for insights.

**Author Contributions** J.A.S., A.B.S., S.N., M.T.M., B.V. and J.V. designed the experiments. S.N., J.V., A.B.S., A.P.R., B.V., M.T.M., R.E.T., E.H. and R.S. carried out the analysis; J.A.S., J.V., A.B.S., S.N., A.P.R. and S.J. wrote the paper; and all other authors carried out various aspects of experimental data collection.

**Author Information** All genomic DNase I footprinting sequence data are available through the NCBI Gene Expression Omnibus (GEO) data repository (accessions GSE26328 and GSE18927), and also through the UCSC browser under the Digital Genomic Footprinting (DGF) table designation. All other data are available through the ENCODE Consortium data release website (see Data Downloads in Supplementary Methods for URL). Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and the online version of the paper is freely available to all readers. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.A.S. (jstam@uw.edu).

doi:10.1038/nature11245

# Architecture of the human regulatory network derived from ENCODE data

Mark B. Gerstein[1,2,3]*, Anshul Kundaje[4]*, Manoj Hariharan[5]*, Stephen G. Landt[5]*, Koon-Kiu Yan[1,2]*, Chao Cheng[1,2]*, Xinmeng Jasmine Mu[1]*, Ekta Khurana[1,2]*, Joel Rozowsky[2]*, Roger Alexander[1,2]*, Renqiang Min[1,2,6]*, Pedro Alves[1]*, Alexej Abyzov[1,2], Nick Addleman[5], Nitin Bhardwaj[1,2], Alan P. Boyle[5], Philip Cayting[5], Alexandra Charos[7], David Z. Chen[3], Yong Cheng[5], Declan Clarke[8], Catharine Eastman[5], Ghia Euskirchen[5], Seth Frietze[9], Yao Fu[1], Jason Gertz[10], Fabian Grubert[5], Arif Harmanci[1,2], Preti Jain[10], Maya Kasowski[5], Phil Lacroute[5], Jing Leng[1], Jin Lian[11], Hannah Monahan[7], Henriette O'Geen[12], Zhengqing Ouyang[5], E. Christopher Partridge[10], Dorrelyn Patacsil[5], Florencia Pauli[10], Debasish Raha[7], Lucia Ramirez[5], Timothy E. Reddy[10]†, Brian Reed[7], Minyi Shi[5], Teri Slifer[5], Jing Wang[1], Linfeng Wu[5], Xinqiong Yang[5], Kevin Y. Yip[1,2,13], Gili Zilberman-Schapira[1], Serafim Batzoglou[4], Arend Sidow[14], Peggy J. Farnham[9], Richard M. Myers[10], Sherman M. Weissman[11] & Michael Snyder[5]

**Transcription factors bind in a combinatorial fashion to specify the on-and-off states of genes; the ensemble of these binding events forms a regulatory network, constituting the wiring diagram for a cell. To examine the principles of the human transcriptional regulatory network, we determined the genomic binding information of 119 transcription-related factors in over 450 distinct experiments. We found the combinatorial, co-association of transcription factors to be highly context specific: distinct combinations of factors bind at specific genomic locations. In particular, there are significant differences in the binding proximal and distal to genes. We organized all the transcription factor binding into a hierarchy and integrated it with other genomic information (for example, microRNA regulation), forming a dense meta-network. Factors at different levels have different properties; for instance, top-level transcription factors more strongly influence expression and middle-level ones co-regulate targets to mitigate information-flow bottlenecks. Moreover, these co-regulations give rise to many enriched network motifs (for example, noise-buffering feed-forward loops). Finally, more connected network components are under stronger selection and exhibit a greater degree of allele-specific activity (that is, differential binding to the two parental alleles). The regulatory information obtained in this study will be crucial for interpreting personal genome sequences and understanding basic principles of human biology and disease.**

A central goal in biology is to understand how a limited cohort of transcription factors is able to organize the large diversity of gene-expression patterns in different cell types and conditions. Over the past decade, system-wide analyses of transcription-factor-binding patterns have been performed in unicellular model organisms, such as *Escherichia coli* and yeast, and have revealed a great deal of information about the organization of regulatory information[1–8]. These studies have provided insights into such features as network hubs[1], connectivity correlations[9], hierarchical organization[10,11] and network motifs[12,13]. Moreover, more complex networks that integrate disparate forms of genomic and proteomic data, such as protein–protein interactions and phosphorylation, have related gene regulation to other biological processes[14–16]. However, for humans, systems-level analyses have been a challenge due to the size of the transcription factor repertoire and genome, and only specific regulatory subnetworks with a handful of factors have been reported thus far[17–19]. The large-scale data from the ENCODE project now begins to enable such analyses[20]. Moreover, with the vast amount of human polymorphism data and genome sequences of many mammals[21,22], it is possible to obtain an unprecedented view of how selection relates to networks.

**ENCODE**
Encyclopedia of DNA Elements
nature.com/encode

Here we present an analysis of the genome-wide binding profiles of 119 transcription-related factors, including sequence-specific, general and chromatin-acting factors. (For simplicity, we refer to all of these as transcription factors, and we use TFSS to denote canonical sequence-specific factors.) We first used the transcription-factor-binding data to analyse the co-association patterns between different factors, as well as their differential patterns in promoter-proximal and distal regulatory regions. We then organized the binding patterns into a stratified hierarchy representing the overall systems-level regulatory wiring. To this, we added other forms of network information, including non-coding RNA (ncRNA) regulation (especially microRNAs

[1]Program in Computational Biology and Bioinformatics, Yale University, Bass 432, 266 Whitney Avenue, New Haven, Connecticut 06520, USA. [2]Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Avenue, New Haven, Connecticut 06520, USA. [3]Department of Computer Science, Yale University, 51 Prospect Street, New Haven, Connecticut 06511, USA. [4]Department of Computer Science, Stanford University, 318 Campus Drive, Stanford, California 94305, USA. [5]Department of Genetics, Stanford University, 300 Pasteur Drive, M-344 Stanford, California 94305, USA. [6]Department of Machine Learning, NEC Laboratories America, 4 Independence Way, Princeton, New Jersey 08540, USA. [7]Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, Connecticut 06520, USA. [8]Department of Chemistry, Yale University, 225 Prospect Street, New Haven, Connecticut 06520, USA. [9]Department of Biochemistry and Molecular Biology, University of Southern California, Norris Comprehensive Cancer Center, 1450 Biggy Street, NRT 6503, Los Angeles, California 90089, USA. [10]HudsonAlpha Institute for Biotechnology, 601 Genome Way, Huntsville, Alabama 35806, USA. [11]Department of Genetics, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA. [12]Genome Center, University of California-Davis, 451 Health Sciences Drive, Davis, California 95616, USA. [13]Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong. [14]Department of Pathology, Stanford University, SUMC L235 (Edwards Bldg), 300 Pasteur Drive, Stanford, California 94305, USA. †Present address: Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, North Carolina 27710, USA.

*These authors contributed equally to this work.

(miRNAs))[23,24], protein–protein interactions[25,26], and protein phosphorylation[27]. We analysed this 'meta-network' for properties that differ based on hierarchical level and connectivity (for example, hubs versus non-hubs) and also searched for enriched network motifs. Finally, we surveyed the pattern of sequence variation over the network, examining selective pressure and allelic effects (preferential binding to the maternal or paternal allele). Several of our key findings are summarized below.

- Human transcription factors co-associate in a combinatorial and context-specific fashion; different combinations of factors bind near different targets, and the binding of one factor often affects the preferred binding partners of others. Moreover, transcription factors often show different co-association patterns in gene-proximal and distal regions.
- Different parts of the hierarchical transcription factor network exhibit distinct properties. For instance, the middle level has the most information-flow bottlenecks and, offsetting this, tends to have the most regulatory collaboration between transcription factors. Conversely, higher-level transcription factors have the greatest connectivity with other networks (for example, the phosphorylome).
- The occurrence of the feed-forward loops is strongly enriched in the transcription factor network, as are a number of motifs in which two genes co-regulated by a factor are bridged by a protein–protein interaction or regulating miRNA.
- Highly connected network elements (both transcription factors and targets) are under strong evolutionary selection and exhibit stronger allele-specific activity (this is particularly apparent when multiple factors are involved). Surprisingly, however, elements with allelic activity are under weaker selection than non-allelic ones.

## Overview of data and processing

The ENCODE project has generated chromatin immunoprecipitation and high-throughput sequencing (ChIP-seq) data sets for 119 distinct transcription factors over five main cell lines (Supplementary Information, section B.1, and Supplementary Tables 1 and 2a). Each data set contains at least two biological replicates. In addition, for a select set of factors (Supplementary Fig. 1c), short interfering RNA (siRNA) experiments were performed, where the transcription factor was depleted and expression changes were quantified by RNA-seq (Supplementary Information, section B.2). Most of the factors (88, 74%) are TFSSs that can be subcategorized on the basis of their DNA-binding domain sequences (Supplementary Table 2a)[28]. A small subset (16, 13%) comprises POL2 and general transcriptional machinery; a final subset (15, 13%) consists of chromatin-modifying and remodelling factors.

To allow effective integrative analysis of these diverse data sets, we developed a uniform processing pipeline and quality-control measures (Supplementary Information, section B.1, and Supplementary Figs 1a, b and 2a; data at http://www.encodeproject.org). In total, we identified 7,424,765 peaks; 2,948,387 (~40%) were proximal (within ±2.5 kilobases) to annotated gene transcription start sites (TSSs).

## Context-specific transcription factor co-association

We first examined the genome-wide co-association of all pairs of transcription factors by analysing the overlap between peaks of all pairs of factors[20]. Although many general trends can be identified, this approach does not take into account the context-specificity of transcription factor binding (that is, the observation that factors bind together in distinct combinations at different genomic locations, and that the co-binding of one pair of transcription factors is often affected by the binding of another transcription factor; Supplementary Information, section C.1). Therefore, we developed a framework focusing on the specific genomic regions bound by a particular transcription factor (the focus factor) and examined the co-association of all other factors (partner factors) within this context (Supplementary Fig. 2a). For each ~350-base-pair region in the focus-factor context, we extracted normalized binding signals of overlapping peaks of all transcription factors, generating a co-binding map. Figure 1a shows such a map for the GATA1 context. Here, factors that consistently co-associate with each other and a substantial proportion of GATA1 peaks are termed 'primary partners' (for example, group 6 transcription factors such as GATA2 and TAL1 in Fig. 1a). In addition to these factors, there are also groups of 'local partners' that co-associate with each other in the presence of GATA1, but only at specific subsets of GATA1-binding peaks (for example, JUN in group 7 and MAX in group 3; Fig. 1a and Supplementary Fig. 2c-1). These 'biclusters', typically containing two to five transcription factors, can be mutually exclusive or partially overlapping.

To identify systematically all primary and local partners for each focus-factor context, we used a machine-learning approach. We derived nonlinear, combinatorial models of each focus-factor's co-binding map relative to randomized control maps (Supplementary Information, section C.2, and Supplementary Fig. 2a, b). Analysis of multivariate rules in these models, in turn, identified pairs and higher-order clusters of significantly co-associated transcription factors. Moreover, these co-associations are robust to peak overlap and calling thresholds (Supplementary Information, section C.4).

The first statistic derived from the models is a relative importance (RI) score (Supplementary Information, section C.2.4.2), which gives the overall importance of each transcription factor in the model. It reflects the 'size' of the biclusters to which a particular transcription factor belongs, and it is related to the number of co-binding factors and the fraction of peak locations involved. For the GATA1 context (Fig. 1b and Supplementary Fig. 2c-2), primary partners TAL1, GATA2 and POL2, as well as local partners MAX and JUN, have high RI scores. To reveal further the partnering in the focus-factor context, we computed co-association scores between all pairs and higher-order sets of transcription factors (Supplementary Information, section C.2.4). These scores measure the impact of the co-dependency implicit in a particular pair on the model as a whole, and they more directly probe the co-occupancy of transcription factors in the focus-factor context than does the RI score. For the GATA1 context, the co-association scores revealed both expected and novel pairings (for example, MYC–MAX–E2F6 and CCNT2–HMGN3, respectively; Fig. 1b, Supplementary Fig. 2c-2 and Supplementary Information, section C.3.1.4). Furthermore, GATA1 is usually associated with enhancer activity. However, the co-association score shows that it is connected to both repressive (for example, NRSF (also called REST) and HDAC2) and activating factors (for example, P300). This discordant behaviour has been observed previously[29]; here, it is borne out by expression studies and knockdowns (Supplementary Information, section C.3.1.4). In particular, after GATA1 knockdown, we found that 94 targets of GATA1 were significantly upregulated, and only 54 were downregulated (Supplementary Fig. 2e-4). Finally, we analysed the functions of genes that lie near clusters of co-associated factors, and found that many are enriched for specific biological functions (Supplementary Fig. 2e-2). For example, one bicluster involving E2F6 (E2F6–GATA1–GATA2–TAL1) was enriched for genes related to myeloid differentiation, whereas another (E2F6–SP1–SP2–FOS–IRF1) was involved in DNA damage response (Supplementary Information, section C.3.3). Thus, distinct combinations of factors regulate specific types of genes.

## Comparing co-association across contexts

### Aggregate RIM and PPM

After establishing the co-binding structure in each transcription factor context, we compared our co-association statistics across contexts. In particular, we combined the RI scores for each transcription factor into a single matrix (RIM, Supplementary Fig. 2a). Clustering reveals nine functionally distinct classes of transcription factor contexts that fall into four broad groups: proximal, distal, repressive

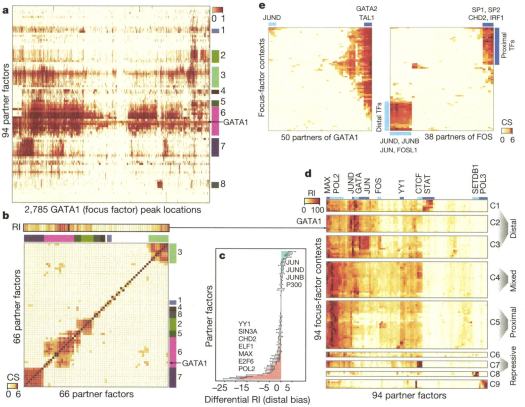


**Figure 1 | Transcription factor co-association. a**, The co-binding map for the GATA1 focus-factor context in K562 cells shows the binding intensity of peaks of all transcription factors (TFs) in K562 (rows) that overlap each GATA1 peak (columns). The coloured rectangles represent eight key clusters consisting of different combinations of co-associating partner-factors. **b**, The GATA1 context-specific relative importance (RI) scores of all partner factors (top) and the matrix of co-association scores (CS) between all pairs of factors (bottom). Primary and local partners of GATA have high RI scores. The co-association score matrix captures the eight clusters observed in **a**. **c**, Different partner factors are preferentially enriched at gene-distal (positive differential RI) and proximal (negative differential RI) GATA1 peaks. **d**, The aggregate factor importance matrix (RIM), obtained by stacking the RI scores of all partner factors (columns) from all focus-factor contexts (rows) in K562 cells, shows nine functionally distinct clusters (C1 to C9) of contexts that can be broadly grouped as distal, proximal, mixed and repressive. The blue rectangles highlight representative partner factors with high RI scores in the clusters. The arrow from **b** to **d** indicates that the GATA1 context-specific RI scores form one row in this matrix. **e**, Co-association variability map of partners (columns) of GATA1 (left panel) and FOS (right panel) over all K562 focus-factor contexts (rows). TAL1 and GATA2 show consistently high co-association scores with GATA1 over most focus-factor contexts, but JUND shows context-specific co-association. FOS shows marked changes in co-association score of partner factors over different contexts (for example, FOS–JUND in distal contexts and FOS–SP2 in proximal ones). (More details are available in Supplementary Fig. 2c, d, f-1, l-2.)

and mixed (Fig. 1d, Supplementary Fig. 2f-1 and Supplementary Information, section C.3.4.1). Next, combining the co-association scores from all focus factors across different contexts provides an overall view of all the primary partners of each transcription factor in the form of a primary-partner matrix (PPM; Supplementary Fig. 2f-4). The RIM reflects the overall similarities in the binding context of focus factors, whereas the PPM highlights the specific factors that tend to co-bind with each other (mutual primary partners). To some degree, one can see the PPM as a subset of the relationships implicit in the RIM. That is, two factors can have similar binding contexts without explicit co-association—for example, two factors that tend both to bind promoters but near different sets of genes. Overall, the PPM shows well known sets of co-associated transcription factors, such as FOS–JUN (the AP1 complex[30,31]) and CTCF–RAD21–SMC3 (the cohesion complex[32,33]), as well as many novel co-associations, such as CHD2–ZBTB33, EGR1–ZBTB7A and CTCF–ZNF143–SIX5 (Supplementary Information, section C3.6.2). We confirmed one novel co-association (CEBPB–TAL1) using co-immunoprecipitation and mass spectrometry (Supplementary Table 3a).

## Variability map

The variability map shows the degree of variability in the partners of a given transcription factor over contexts (as determined by the co-association score) (Supplementary Information, section C.2.5.5). For instance, Fig. 1e shows that GATA1 has mostly the same partners in many contexts (for example, TAL1 and GATA2 are partners over almost all contexts). However, a few partners (for example, JUND) are present in only some contexts. An extreme example is FOS, which completely changes its partners in different contexts (Fig. 1e, Supplementary Fig. 2l-2 and Supplementary Information, section C.3.6.1).

## Cell-type differences

We analysed transcription factor co-association in the five main ENCODE cell types (Supplementary Information, section C.3.4). The GM12878 and K562 cell lines have the most common (31) transcription factor data sets (Supplementary Information, section C.3.5). Comparative analysis showed that over 80% of the transcription factor pairs had no significant change in co-association between K562 and GM12878 cell lines. However, there were a few marked examples of cell-line differences. For instance, FOS and JUND co-associate in K562 but not in GM12878 cells (Supplementary Information, section 3.5.1), despite the fact that most of the other partners of FOS are maintained in both cell lines.

## Gene context: proximal versus distal

Overall, we found distinct partner preferences at proximal and distal sites. These results were robust to the choice of the distance used to define proximal and distal regions (Supplementary Fig. 2c-3). In particular, for the GATA1 context, we found that RI scores change markedly between proximal or distal sites (Fig. 1c and Supplementary Fig. 2c-3): typical core promoter transcription factors (for example, POL2, E2F6, MAX and ELF1) have a significant proximal promoter bias, whereas JUND, JUNB, JUN and P300 show preferential co-association with distal sites. Another way of analysing differences between proximal and distal sites is in the framework of the variability map, in which one can observe the changing partners of a transcription factor in different contexts. For instance, FOS has completely different partners with which it co-associates proximally and distally (Fig. 1e, Supplementary Fig. 2l-2 and Supplementary Information, section C.3.6.1).

## Assembling pairwise interactions into hierarchies

Analysis of co-associations specifies the relationships between the DNA-binding profiles of multiple regulators. To obtain a systems-level perspective, we recast transcription factor associations as a network (Supplementary Fig. 4a), wherein the nodes are regulators or their targets, and the edges designate regulatory relationships. Here, we focussed on the global wiring pattern across all cell types. We expected different subnetworks within this framework to be active to different degrees in different cells.

Using our binding-site list, we identified an initial set of regulatory targets from genes having promoter-proximal binding sites. The resulting raw network consists of 500,542 promoter-associated interactions between transcription factors and all their putative targets, of which 4,809 are between pairs of factors (networks at http://encodenets.gersteinlab.org). We filtered this to identify the most confident interactions using a probabilistic model, giving 26,070 total interactions, with only 338 between transcription factors[34] (Supplementary Information, section D.1). We validated the performance of the filtering using the siRNA experiments; for each case, the targets identified by our model were more differentially expressed in siRNA-treated cells than were those identified by a simple peak-based method (Supplementary Fig. 1c–e).

We next computed common connectivity statistics for individual transcription factors, namely, out-degree ($O$), in-degree ($I$) and betweenness, which were then used to identify hubs and information-flow bottlenecks (Supplementary Information, section K). Of particular interest is the difference between out- and in degree ($O - I$), which measures the direction of information flow (Supplementary Fig. 3a). A positive value suggests that a transcription factor is located 'upstream' in the network, whereas a negative value indicates that it is 'downstream'. We further defined a normalized version of this 'hierarchy height' metric, $h = (O - I)/(O + I)$. We found that this can be approximated by three levels (Supplementary Fig. 3c), with top-level, 'executive' transcription factors regulating many other factors ($h \approx 1$), and bottom-level 'foreman' factors more regulated than regulating ($h \approx -1$). For purposes of visualization, we used a simulated-annealing procedure to optimally and robustly arrange the 119 transcription factors into three discrete levels (with the number of downward-pointing edges maximized) (Fig. 2a and Supplementary Information, section D.2).

## Layering on distal, ncRNA and protein interactions

The filtered transcription factor hierarchy consists of the strongest promoter-associated interactions. Building upon this skeleton, we added additional types of connections.

Interactions involving distal regulatory elements (for example, enhancers) are more difficult to identify than those involving proximal elements. Here, we used a statistical model[35]. This identifies distal sites with potentially many binding transcription factors using chromatin features. These regions were associated with a gene if their changing pattern of chromatin marks across cell lines correlates with the expression of that gene (Supplementary Information, section E.1). Overall, the model identified 19,258 distal edges (Fig. 2a).

The regulatory interactions between transcription factors and ncRNAs constitute an additional layer of information to add to the meta-network. We used transcription factor peaks proximal to ncRNAs to identify transcription-factor-to-ncRNA regulation. Next, we incorporated miRNA-to-transcription-factor regulatory interactions from TargetScan[36] (Supplementary Information, section E.2). Finally, we incorporated physical protein–protein interactions[26], as well as predicted phosphorylations (Supplementary Information, section F.3, and Supplementary Fig. 7a). Overall, these different interactions form a dense meta-network that we analysed further for interesting biological properties.

## Relating network connectivity and genomic properties

We next correlated measures for the connectivity and hierarchical position of each transcription factor with a wide variety of genomic and proteomic properties (Fig. 2c, Table 1 and Supplementary Table 4, $P$ values in the latter).

### Correlations with distal edges

Distal edges have a different degree distribution than do proximal ones (Fig. 2a and Supplementary Fig. 5). Inspection reveals that many point upward in the transcription factor hierarchy, opposite to most proximal edges. Furthermore, we found many transcription factors with low in-degree values in the proximal network but high in-degree values in the distal one, indicating that they are heavily regulated through enhancers (Supplementary Fig. 5a). Some of these are well known condition- and tissue-specific regulators (for example, IRF4 and GATA1)[37].

### Correlations within the proximal network

Upper-level transcription factors tend to have more targets than lower-level ones, both overall and when considering only other transcription factors as targets. As measured by betweenness in proximal regulation, middle-level transcription factors form information-flow bottlenecks (Fig. 2c). Moreover, betweenness in the proximal network is correlated with more distal regulation. This tends to increase the information flow through mid-level bottlenecks even more. (See Supplementary Information section F.3.6 for clarification of the implications.)

### Correlation with protein interactions and the phosphorylome

We found that top-level transcription factors tend to have more partners in the protein–interaction network than do lower-level ones (Fig. 2c and Table 1). We further studied how transcription factors in different levels are regulated by kinases. Although there is no significant difference in terms of the number of kinases regulating transcription factors at different levels, we found that if the phosphorylome is arranged into a hierarchy using the same approach used for organizing the transcription factor network, kinases at the bottom tend not to phosphorylate transcription factors, but they tend to be regulated by them (particularly by top-level factors; Supplementary Fig. 7).

### Correlation with ncRNAs

We found that top- and middle-level transcription factors have the highest total number of ncRNA targets (Fig. 2c, Table 1 and Supplementary Fig. 6a), consistent with our findings for protein-coding targets. We then developed a score indicating the fraction of a transcription factor's total regulation devoted to ncRNAs, relative to protein-coding genes (Supplementary Information, section E.2); this identified several factors that preferentially target ncRNAs, such as BDP1 and BRF2 (Supplementary Fig. 6b, c).

Matching the pattern for ncRNAs in general, most of the transcription factors involved in miRNA regulation tend to be top- or middle-level ones (Fig. 2c). Moreover, highly connected transcription factors tend to regulate more miRNAs and to be more regulated by them

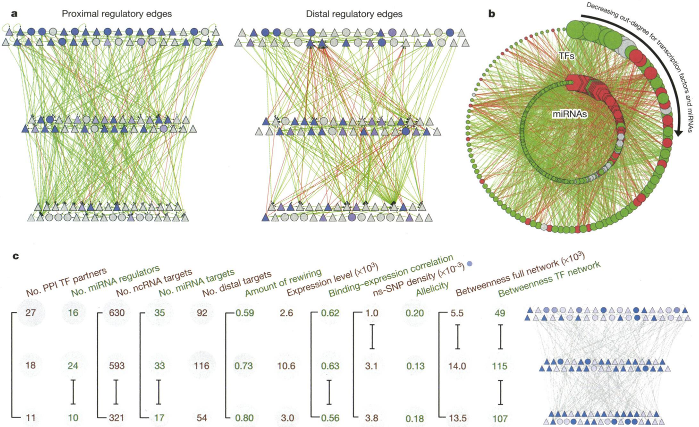

a
Proximal regulatory edges
Distal regulatory edges
b
Decreasing out-degree for transcription factors and miRNAs
TFs
miRNAs
c
No. PPI TF partners
No. miRNA regulators
No. ncRNA targets
No. miRNA targets
No. distal targets
Amount of rewiring
Expression level (×10³)
Binding–expression correlation
ns-SNP density (×10⁻³)
Allelicity
Betweenness full network (×10³)
Betweenness TF network
27 16 630 35 92 0.59 2.6 0.62 1.0 0.20 5.5 49
18 24 593 33 116 0.73 10.6 0.63 3.1 0.13 14.0 115
11 10 321 17 54 0.80 3.0 0.56 3.8 0.18 13.5 107

**Figure 2 | Overall network.** **a**, Close-up representation of the transcription factor hierarchy. Nodes depict transcription factors. TFSSs are triangles, and non-TFSSs are circles. Left: proximal-edge hierarchy with downward pointing edges coloured in green and upward pointing ones coloured in red. The nodes are shaded according to their out-degree in the full network (as described in Table 1). Right: factors placed in the same proximal hierarchy but now with edges corresponding to distal regulation coloured green and red, and nodes re-coloured according to out-degree in the distal network. The distal edges do not follow the proximal-edge hierarchy. **b**, Close-up view of transcription-factor–miRNA regulation. The outer circle contains the 119 transcription factor, whereas the inner circle contains miRNAs. Red edges correspond to miRNAs regulating transcription factors; green edges show transcription factors regulating miRNAs. Transcription factors and miRNAs each are arranged by their out-degree, beginning at the top (12:00) and decreasing in order clockwise. Node sizes are proportional to out-degree. For transcription factors, the out-degree is as described in Table 1; for miRNAs, it is according to the out-degree in this network. Red nodes are enriched for miRNA–transcription factor edges and green nodes are enriched for transcription factor–miRNA edges. Grey nodes have a balanced number of edges (within ±1). **c**, Average values of various properties (topological, dynamic, expression-related and selection-related—ordered consistently with Table 1) for each level are shown for the proximal-edge hierarchy. The top, middle and bottom rows correspond to the top, middle and bottom of the hierarchy, respectively. The sizing of the grey circles indicates the relative ordering of the values for the three levels. Significantly different values ($P < 0.05$) using the Wilcoxon rank-sum test are indicated by black brackets. The proximal-edge hierarchy depicted on the right shows non-synonymous SNP (ns-SNP) density, where the shading corresponds to the density for the associated factor. (See Supplementary Fig. 4 for more details.)

**Table 1 | Correlating properties with centrality and hierarchy height**

| Category | Property | Correlation with: | | | | |
|---|---|---|---|---|---|---|
| | | Degree centrality‡ | | Betweenness centrality | | (O − I)/(O + I) |
| | | Full | TF–TF | Full | TF–TF | TF–TF |
| Topology | Number of TF partners in PPI | 0.28† | 0.27† | 0.25* | 0.33† | 0.08 |
| Topology | Number of miRNA regulators | 0.24* | 0.33† | −0.02 | 0.00 | 0.29† |
| Topology | Number of ncRNA targets | 0.65† | 0.49† | 0.34† | 0.35† | 0.22* |
| Topology | Number of miRNA targets | 0.62† | 0.50† | 0.33† | 0.34† | 0.19* |
| Topology | Number of distal targets | 0.32† | 0.24* | 0.19* | 0.23* | 0.07 |
| Dynamics | Amount of rewiring | −0.14 | −0.12 | 0.44* | 0.35 | −0.42* |
| Expression | Expression level | 0.14 | 0.12 | 0.23* | 0.27* | −0.04 |
| Expression | Binding–expression correlation | 0.41† | 0.31† | 0.30† | 0.36† | 0.19* |
| Selection properties for factors | ns-SNP density | −0.19* | −0.27* | −0.01 | −0.03 | −0.22 |
| Selection properties for factors | Allelicity | 0.20 | 0.28* | −0.10 | −0.16 | 0.18 |
| Selection properties for targets | ns-SNP density | −0.05† | – | – | – | – |
| Selection properties for targets | dN/dS | −0.05† | – | – | – | – |

Spearman correlation values of various properties (topological, dynamic, expression-related and selection-related) with centrality measures and hierarchy height. Only properties that are significantly correlated with centrality or hierarchy height are listed. For a full set of properties, *P* values and explanations, see Supplementary Tables 4 and 6. dN/dS, non-synonymous to synonymous mutation ratio.

* Spearman correlation $P < 0.05$.

† Spearman correlation $P < 0.01$.

‡ Degree centrality refers to out-degree, except for selection properties on targets, in which case it refers to in-degree. In particular, out-degree in the full transcription factor target network refers to the 'Targets' column in Supplementary Table 4a, and the same quantity is used throughout Fig. 2.

(Table 1 and Fig. 2b). However, when we analyse transcription-factor–miRNA regulation in detail we find that the factors most involved in miRNA regulation tend to either largely regulate or be regulated by miRNAs (Fig. 2b and Supplementary Fig. 4d). That is, there are few high-degree transcription factors with 'balanced regulation' (similar numbers of incoming and outgoing edges, relative to a control; Supplementary Fig. 3m). The same pattern can be seen for miRNAs (Supplementary Fig. 3l).

### Correlation with families and functional categories

Chromatin-related factors are enriched at the top of the hierarchy, whereas TFSSs are enriched in the middle (Supplementary Table 5a and Supplementary Information, section F.1). Also, TFSSs show a greater degree of tissue specificity and are more highly regulated by miRNAs than are general and chromatin-related factors (Supplementary Information, section F.4), indicating that they may be more finely tuned in their expression. Examining functional enrichment, we found that transcription factors at the top of the hierarchy tend to have more general functions, and those at the bottom tend to have more specific functions (Supplementary Table 5c and Supplementary Information, section F.1).

### Correlation with network dynamics

We studied how transcription factors change their binding patterns among different cell types, principally between the K562 and GM12878 cell lines. We quantified the amount of 'rewiring' as the fraction of unshared targets, normalized by the union of two target sets (Supplementary Information, section 3.5). We found that this 'rewiring score' is negatively correlated with hierarchy height (Fig. 2c and Table 1). This means that the targets of lower-level transcription factors tend to change more between cell types, consistent with their role in more specialized processes.

### Correlation with gene expression

We calculated the average expression levels of transcription factors across 34 tissues[26]; highly connected factors tend to be highly expressed. We further examined the relationship between connectivity and expression by calculating, for each transcription factor, the correlation between its binding signal around its targets and the level of target expression (Supplementary Information, section F.3.4). This binding–expression correlation is positively correlated with factor connectivity. Moreover, transcription factors at the top and middle levels show a greater correlation. Thus, more 'influential' transcription factors tend to be better connected and higher in the hierarchy. (This degree of 'influence" becomes even clearer when one considers weighting the correlation by the number of transcription factor targets, given that higher-level factors tend to have more targets.) However, somewhat surprisingly, a model integrating the binding–expression relationships of all the highly connected transcription factors has about the same predictive power for expression as a model integrating all the less connected ones, indicating that the weak binding–expression relationships of the less influential factors are collectively quite influential (Supplementary Information, section F.3.4)[38].

## Collaboration between hierarchy levels

We explored how transcription factors in the top, middle and bottom (T, M and B, respectively) levels of the hierarchy collaborate, in terms of both inter-level (TM, MB, TB) and intra-level (TT, MM, BB) relationships (Fig. 3a). We examined three kinds of collaboration: co-association (as described earlier), physical interactions, and target-expression cooperativity. We defined two transcription factors as being cooperative if their shared targets are significantly different in expression from their unshared targets (Supplementary Information, section G.2). Overall, we found that collaborations involving the middle level (and to a lesser extent, the top one) tended to be enriched. In particular, TM and MM transcription factor pairs influenced gene expression cooperatively. Next, all co-associations involving top- and middle-level factors are enriched, whereas those involving the bottom level are depleted. A similar pattern was observed for protein–protein interactions, with TT and TM co-regulation more likely to occur between physically interacting transcription factors (Fig. 3a and Supplementary Information, section G.1).

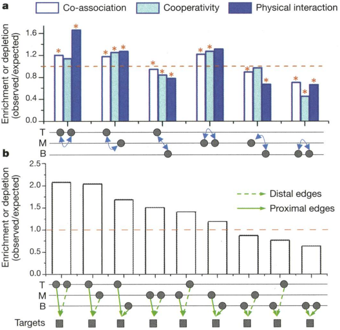


**Figure 3 | Collaboration between levels.** **a**, Enrichment of collaborating transcription factor pairs from different levels (top (T), middle (M) and bottom (B)). The factors are represented by two nodes below each bar graph. The dashed orange line indicates the expected level of collaboration. Significant enrichment above or depletion below that level is marked by asterisks ($P < 0.05$). (See Supplementary Information section G.1.2 for more details.) **b**, Enrichment of proximal and distal co-regulatory pairs in the network hierarchy. Co-regulatory pairs from different levels are shown by the two nodes below each bar.

Finally, we analysed how proximal and distal sites 'collaborate'. We identified pairs of transcription factors that bind to the promoter and distal regulatory regions of the same target gene (Supplementary Information, section G.3) and studied their respective locations in the factor hierarchy. We found an asymmetry between proximal and distal regulation, with transcription factors associated through promoter regulation more likely to reside in upper levels (Fig. 3b).

## Enriched network motifs

Apart from its global structure, we further studied the network from the perspective of its constituent building blocks; that is, network motifs, which are small connectivity patterns that carry out canonical functions[39]. We systematically searched for motifs, first in the promoter-regulation hierarchy and then in the meta-network including distal, miRNA and protein–protein interactions. Our procedure was to instantiate all possible motifs for broad template patterns and then determine which of these were significantly over- or under-represented relative to a random control[40] (Supplementary Information, section H). For instance, starting with all possible three-transcription-factor motifs in the proximal network (Fig. 4a), we found the most enriched motif to be the well-studied feed-forward loop (FFL)[39]. In agreement with the observed collaborations within the hierarchy, many FFLs involve the middle level (Supplementary Fig. 9a). Moreover, by analysing the expression levels of the constituent genes of the FFLs over many tissues, we found that many were positively correlated, highlighting the tight regulation implicit in the motif (Fig. 4a and Supplementary Information, section H.1).

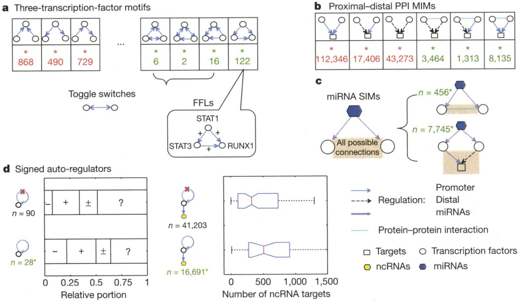


**Figure 4 | Motif analysis.** Motifs are accompanied by the number of occurrences, *n*. Enriched motifs are highlighted in green; depleted ones in red. An asterisk means that the corresponding enrichment/depletion is statistically significant ($P = 1 \times 10^{-5}$). The motifs are sorted such that those at the ends have more significant *P* values. (See Supplementary Fig. 9h for more details.) **a**, Systematic search of three-transcription-factor motifs. The most enriched motif is the FFL. A particular example formed by STAT1, STAT3 and RUNX1 is highlighted. Here, the '+' symbol on an edge indicates that the correlation between the gene expression of the source and the target across tissues is positive. Other motifs containing a toggle-switch regulation on top of the basic FFL design are also indicated. **b**, Proximal–distal PPI MIMs. Here we searched all motifs involving the co-regulation of two transcription factors (which could be either proximal or distal) with (or without) a protein–protein interaction between them. Motifs containing the protein–protein interaction tended to be enriched. **c**, miRNA SIMs. The two enriched motifs resulting from enumerating all motifs in which a miRNA targets two transcription factors that are connected in various ways are shown. These two motifs contain a protein complex of two transcription factors and a cooperative pair of promoter and distal regulatory transcription factors. **d**, The auto-regulator motif is enriched in the transcription factor–transcription factor network: 28 of all factors are auto-regulators. Moreover, auto-regulators are more likely to be repressors (−) relative to non-auto regulators, and they tend to have more ncRNAs as their targets. In the box plots, the red line indicates the median, the blue box shows the interquartile range (IQR), and whiskers extend out to 1.5 IQR.

Finally, we found further enriched three-transcription-factor motifs containing an additional regulation on top of that in a FFL. This creates a mutual regulation between a pair of transcription factors, instantiating a toggle-switch, which has been shown to have an essential role in the determination of cell fate[41].

Next, we analysed another template: all possible multiple-input modules (MIMs, defined in Supplementary Information, section K) involving promoter and distal regulation and a protein–protein interaction (proximal–distal PPI MIMs, Fig. 4b). We found that co-regulating transcription factors are likely to interact physically, indicating that they work together as a complex. Moreover, the motif ranking second in enrichment consists of a distal regulatory relationship, a promoter regulatory relationship, and a protein–protein interaction. This is suggestive of a common picture of DNA looping, with an interacting complex of transcription factors binding to the promoter and enhancer simultaneously.

The connection between co-regulated entities extends to miRNA regulation. We surveyed all possible instances of a miRNA regulating two transcription factors (miRNA SIM, Fig. 4c) and found that the miRNAs are more likely to regulate a pair of physically interacting factors. This enrichment indicates that, to avoid unwanted cross-talk, a miRNA tends to shut down an entire functional unit (that is, transcription factor complex) rather than just a single component. Similarly, we found that miRNAs tend to target a pair of transcription factors binding both proximally and distally (Fig. 4c). This suggests that miRNA represses the expression of both promoter and distal regulators to shut down a target completely. Apart from miRNAs, we also studied motifs involving other kinds of ncRNAs. Among motifs involving a transcription factor regulating two ncRNAs, there is great enrichment for both ncRNAs to be long intergenic noncoding RNAs (lincRNAs) (Supplementary Information, section H.2).

Finally, we found the network to be enriched for auto-regulators (28 out of 119 transcription factors), a simple but important motif, which are commonly found in networks exhibiting multistability[42]. Moreover, we found that the auto-regulators tend to be repressors, representing a well known design principle for maintaining steady state[39] (Fig. 4d).

## Allelic behaviour in a network framework

We examined the relationship between sequence variation and transcription factor regulation. In particular, we investigated the coordination between allele-specific binding and allele-specific expression[43,44]. We used the sequenced data sets for the GM12878 cell line, which has a deeply sequenced diploid genome (Supplementary Information, section I.1). We extended pairwise analysis of allele-specific behaviour[20] to study higher-order coordination of multiple factors regulating a common target. We first generated the unfiltered, promoter-regulation network for GM12878 cells and then identified a subnetwork within it representing the difference between maternal- and paternal-specific networks (Supplementary Information, section I.2). This subnetwork is shown in Fig. 5a, with 4,798 transcription-factor-target edges coloured red or blue to represent predominantly maternally or paternally regulated targets; the targets are similarly coloured to indicate predominantly maternal or paternal expression. We found that of the 4,798 allele-specific binding cases of a single factor regulating its associated target, 57% showed coordinated allelic binding and expression. We then found that for the cases in which two transcription factors regulate a common target, 63% were consistent (that is, both factors bind to the same allele that is expressed). For those cases in which triplets of transcription factors regulate a common target, the consistency increased to 65%. This trend continues, demonstrating that, as one increases the degree of combinatorial regulation, there is a progressively stronger relationship between expressed and regulated alleles.

The degree of allele-specific behaviour of each transcription factor can be quantified by a statistic that we call 'allelicity'. The allelicity of a transcription factor is defined as the fraction of single nucleotide

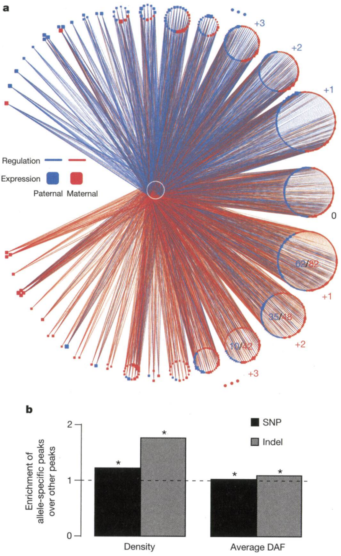


**Figure 5 | Allelic effects. a**, An 'allelic effects network' depicting the increasing coordination between allele-specific binding and allele-specific expression as the number of factors regulating a target increases. Central white nodes denote transcription factors, and peripheral nodes denote targets, which are blue (red) if they are expressed from the paternal (maternal) allele. Blue (red) edges denote allele-specific binding to the paternal (maternal) allele. This network represents the strongest differences between the paternal- and maternal-specific regulatory networks. As one goes around the larger circle anticlockwise (clockwise), each of the small circular clusters represents targets with progressively more paternal (maternal) regulation, indicated by the small blue (red) numbers to the side of the clusters. Moreover, within each of the clusters the fraction of predominantly paternally (maternally) expressed targets increases as one goes around the larger circle. As an illustration, this fraction is explicitly indicated by the ratios within three of the larger clusters at the bottom right. **b**, Relationship between transcription factor allelicity and selection. The bar height is the ratio of the degree of selection (as measured by SNP density or average DAF) in those binding peaks showing allelic behaviour to the degree of selection in all other binding peaks. Asterisks represent significant differences ($P < 0.05$, Wilcoxon rank-sum test). (See Supplementary Information section I.2 and Supplementary Fig. 10b, c for details.)

polymorphisms (SNPs) that exhibit allele-specific binding out of all the SNPs that may potentially exhibit it (Supplementary Information, section I.3). Thus, qualitatively, allelicity may be thought of as the sensitivity of a transcription factor's binding to maternal-versus-paternal variants. Using our network described here, we find that transcription factors with higher degrees of allelicity tend to have more target genes, indicating that these factors tend to vary more in their binding with sequence (Table 1). Finally, we found that small insertions and deletions (indels) tended to cause disproportionally more of these allelic events than did SNPs (Supplementary Table 6g).

## Selection in a network context

Previous studies have examined the relationship between evolutionary selection and position in the human protein–protein interaction network[45]. However, the analogous relationship in the regulatory network has not yet been explored.

### Selection

To address this, we first analysed the selective pressure on both transcription factors and their targets. We predominantly used non-synonymous SNP density from the 1000 Genomes Pilot[21] to determine selection among modern-day humans (Supplementary Information, section J). We also verified our results using other measures of selection (that is, derived allele frequency (DAF) and the ratio of non-synonymous to synonymous SNP rates (pN/pS statistic) (Supplementary Information, section J)). For selection over longer timescales, we calculated the ratio of non-synonymous to synonymous substitution rates in human–chimp orthologue alignments (dN/dS). We found significant negative correlation between the regulatory in-degree of target genes and both their non-synonymous SNP density and dN/dS values (Table 1 and Supplementary Table 6e). Thus, target genes regulated by more transcription factors are under stronger negative selection. Similarly, we found that there is a significant negative correlation between transcription factor regulatory out-degree and non-synonymous SNP density (Table 1 and Supplementary Table 6d). We observed a consistent result with transcription factor dN/dS values and other measures of selection, although these are not all as statistically significant (Supplementary Table 6d and Supplementary Information, section J). This shows that transcription factors regulating more targets tend to be under stronger negative selection. Moreover, within the transcription factor hierarchy, we found that factors at the top are under significantly stronger negative selection (Fig. 2c, Table 1 and Supplementary Table 6b).

Consistent with all of these results relating connectivity with constraint, we found that genes tolerant of loss-of-function mutations[46], which are under weaker negative selection, have a significantly lower total degree ($I + O$) than other genes (Supplementary Information, section J).

### Selection and allelic effects

Finally, we attempted to relate selection and allelic effects. We extracted transcription-factor-binding peaks in promoters and gene bodies showing allele-specific binding, and compared the selective pressure in these against a control (binding peaks within the same regions without allele-specific binding). We found that transcription-factor-binding peaks exhibiting allelic effects have higher SNP densities relative to the control (Fig. 5b). Moreover, binding peaks with no allelic effects show a skew in the DAF spectrum towards rarer SNPs, relative to allele-specific binding ones (Fig. 5b and Supplementary Fig. 10c). The same trend holds true for indels and structural variants (Fig. 5b and Supplementary Fig. 10b, c). Interestingly, these results indicate that allelic regulation seems to be under less selective constraint.

## Discussion

This study provides the first detailed analysis of how human regulatory information is organized. A number of clear design principles emerge from it. Many of these are shared with model organisms (Supplementary Table 7), demonstrating that they are general features of transcription factor regulation. First, we found that the connectivity and hierarchical organization of regulatory factors is reflected in many genomic properties. For instance, top-level transcription factors have their binding more strongly correlated with the expression of their targets, perhaps indicating that they are more influential, as reported for model organisms[47]. Next, the middle-level contains information-flow bottlenecks and much connectivity with miRNA and distal regulation. Targeting these bottlenecks (for example, by drugs) is likely to most strongly affect the flow of information through regulatory circuits. To some degree, the cell mitigates the effect of bottlenecks by having pairs of middle-level transcription factors collaborate in regulation. (Co-regulation mitigates bottlenecks.) Third, the regulatory network seems to be built from repeated reuse of small, modular motifs. In particular, regulation between levels involves many feed-forward loops, which could be used to filter fluctuations in input stimuli. Again, these properties are shared with model organisms; the network motifs and cooperating middle-level have been observed in yeast[48].

By contrast, the differences in proximal and distal regulation seem to be a unique feature of human regulation. This finding is evident in the analysis of both transcription factor co-association and network structure. The proximal–distal differences reflect the much larger intergenic space in humans than model organisms and the commensurately larger amount of distal binding. Finally, analysis of conservation indicates that more highly connected parts of the network are under stronger selection, consistent with results from model organisms. However, one unique finding for humans is 'allelic' effects. More highly connected transcription factors are more likely to exhibit allele-specific binding. Interestingly, we found that the actual allele-specific binding sites tend to be under less selection. Unravelling this interaction between selection and regulatory networks will be crucial to interpreting variants in the many personal genome sequences expected in the future. Co-published ENCODE-related papers can be explored online via the *Nature* ENCODE explorer (http://www.nature.com/ENCODE), a specially designed visualization tool that allows users to access the linked papers and investigate topics that are discussed in multiple papers via thematically organized threads.

## METHODS SUMMARY

Detailed methods associated with each section of the paper are in a similarly titled section of the Supplementary Information. In particular, an overview of our data processing pipeline is in Supplementary Information, section B.

**Received 9 December 2011; accepted 22 May 2012.**

1. Lee, T. I. *et al.* Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298,** 799–804 (2002).
2. Balazsi, G., Barabasi, A. L. & Oltvai, Z. N. Topological units of environmental signal processing in the transcriptional regulatory network of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **102,** 7841–7846 (2005).
3. Yu, H. Y. & Gerstein, M. Genomic analysis of the hierarchical structure of regulatory networks. *Proc. Natl Acad. Sci. USA* **103,** 14724–14731 (2006).
4. Hu, Z. Z., Killion, P. J. & Iyer, V. R. Genetic reconstruction of a functional transcriptional regulatory network. *Nature Genet.* **39,** 683–687 (2007).
5. Balaji, S., Babu, M. M. & Aravind, L. Interplay between network structures, regulatory modes and sensing mechanisms of transcription factors in the transcriptional regulatory network of *E. coli*. *J. Mol. Biol.* **372,** 1108–1122 (2007).
6. Jothi, R. *et al.* Genomic analysis reveals a tight link between transcription factor dynamics and regulatory network architecture. *Mol. Syst. Biol.* **5,** 294 (2009).
7. Barabási, A. L. & Oltvai, Z. N. Network biology: Understanding the cell's functional organization. *Nature Rev. Genet.* **5,** 101–113 (2004).
8. Kim, H. D., Shay, T., O'Shea, E. K. & Regev, A. Transcriptional regulatory circuits: Predicting numbers from alphabets. *Science* **325,** 429–432 (2009).
9. Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Science* **296,** 910–913 (2002).
10. Ma, H. W., Buer, J. & Zeng, A. P. Hierarchical structure and modules in the *Escherichia coli* transcriptional regulatory network revealed by a new top-down approach. *BMC Bioinformatics* **5,** 199 (2004).
11. Balaji, S., Iyer, L. M., Aravind, L. & Babu, M. M. Uncovering a hidden distributed architecture behind scale-free transcriptional regulatory networks. *J. Mol. Biol.* **360,** 204–212 (2006).
12. Milo, R. *et al.* Network motifs: Simple building blocks of complex networks. *Science* **298,** 824–827 (2002).
13. Cosentino Lagomarsino, M., Jona, P., Bassetti, B. & Isambert, H. Hierarchy and feedback in the evolution of the *Escherichia coli* transcription network. *Proc. Natl Acad. Sci. USA* **104,** 5516–5520 (2007).
14. Ptacek, J. *et al.* Global analysis of protein phosphorylation in yeast. *Nature* **438,** 679–684 (2005).
15. Beyer, A., Bandyopadhyay, S. & Ideker, T. Integrating physical and genetic maps: from genomes to interaction networks. *Nature Rev. Genet.* **8,** 699–710 (2007).
16. Yu, H. Y., Xia, Y., Trifonov, V. & Gerstein, M. Design principles of molecular networks revealed by global comparisons and composite motifs. *Genome Biol.* **7,** R55 (2006).
17. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133,** 1106–1117 (2008).
18. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122,** 947–956 (2005).
19. Reed, B. D., Charos, A. E., Szekely, A. M., Weissman, S. M. & Snyder, M. Genome-wide occupancy of SREBP1 and its partners NFY and SP1 reveals novel functional toles and combinatorial regulation of distinct classes of genes. *PLoS Genet.* **4,** e1000133 (2008).
20. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* http://dx.doi.org/10.1038/nature11247 (this issue).
21. Altshuler, D. L. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467,** 1061–1073 (2010).
22. Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478,** 476–482 (2011).
23. Barski, A. *et al.* Chromatin poises miRNA- and protein-coding genes for expression. *Genome Res.* **19,** 1742–1751 (2009).
24. Ozsolak, F. *et al.* Chromatin structure analyses identify miRNA promoters. *Genes Dev.* **22,** 3172–3183 (2008).
25. Stark, C. *et al.* The BioGRID Interaction Database: 2011 update. *Nucleic Acids Res.* **39,** D698–D704 (2011).
26. Ravasi, T. *et al.* An atlas of combinatorial transcriptional regulation in mouse and man. *Cell* **140,** 744–752 (2010).
27. Novershtern, N. *et al.* Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144,** 296–309 (2011).
28. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nature Rev. Genet.* **10,** 252–263 (2009).
29. Kerenyi, M. A. & Orkin, S. H. Networking erythropoiesis. *J. Exp. Med.* **207,** 2537–2541 (2010).
30. Curran, T. & Franza, B. R. Fos and Jun: the AP-1 Connection. *Cell* **55,** 395–397 (1988).
31. Chinenov, Y. & Kerppola, T. K. Close encounters of many kinds: Fos-Jun interactions that mediate transcription regulatory specificity. *Oncogene* **20,** 2438–2452 (2001).
32. Rubio, E. D. *et al.* CTCF physically links cohesin to chromatin. *Proc. Natl Acad. Sci. USA* **105,** 8309–8314 (2008).
33. Parelho, V. *et al.* Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132,** 422–433 (2008).
34. Cheng, C., Min, R. & Gerstein, M. TIP: A probabilistic method for identifying transcription factor target genes from ChIP-Seq binding profiles. *Bioinformatics* **27,** 3221–3227 (2011).
35. Yip, K. Y. *et al.* Classification of human genomic regions based on experimentally-determined binding sites of more than 100 transcription-related factors. *Genome Biol.* **13,** R48 (2012).
36. Friedman, R. C., Farh, K. K. H., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19,** 92–105 (2009).
37. Baron, M. H. & Farrington, S. M. Positive regulators of the lineage-specific transcription factor GATA-1 in differentiating erythroid cells. *Mol. Cell. Biol.* **14,** 3108–3114 (1994).
38. Cheng, C. *et al.* Understanding transcriptional regulation by integrative analysis of transcription factor binding data. *Genome Res.* http://dx.doi.org/10.1101/gr.136838.111 (2012).
39. Alon, U. Network motifs: theory and experimental approaches. *Nature Rev. Genet.* **8,** 450–461 (2007).
40. Cheng, C. *et al.* Construction and analysis of an integrated regulatory network derived from high-throughput sequencing data. *PLoS Comput. Biol.* **7,** e1002190 (2011).
41. Zhou, J. X. & Huang, S. Understanding gene circuits at cell-fate branch points for rational cell reprogramming. *Trends Genet.* **27,** 55–62 (2011).
42. Burda, Z., Krzywicki, A., Martin, O. C. & Zagorski, M. Motifs emerge from function in model gene regulatory networks. *Proc. Natl Acad. Sci. USA* **108,** 17263–17268 (2011).
43. McDaniell, R. *et al.* Heritable individual-specific and allele-specific chromatin signatures in humans. *Science* **328,** 235–239 (2010).
44. Rozowsky, J. *et al.* AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol. Syst. Biol.* **7,** 522 (2011).
45. Kim, P. M., Korbel, J. O. & Gerstein, M. B. Positive selection at the protein network periphery: Evaluation in terms of structural constraints and cellular context. *Proc. Natl Acad. Sci. USA* **104,** 20274–20279 (2007).
46. MacArthur, D. G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335,** 823–828 (2012).
47. Bhardwaj, N., Kim, P. M. & Gerstein, M. B. Rewiring of transcriptional regulatory networks: hierarchy, rather than connectivity, better reflects the importance of regulators. *Sci. Signal.* **3,** ra79 (2010).
48. Bhardwaj, N., Yan, K.-K. & Gerstein, M. B. Analysis of diverse regulatory networks in a hierarchical context shows consistent tendencies for collaboration in the middle levels. *Proc. Natl Acad. Sci. USA* **107,** 6841–6846 (2010).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank the ENCODE Project (National Institutes of Health/National Human Genome Research Institute (NIH/NHGRI)) for funding. We thank P. Bickel and B. Brown for conversations. Funding has also been provided by the NIH Predoctoral Training Program in Biophysics (D.C.; T32 GM008283-24) and the Sackler institute (G.Z.-S.). M.H. thanks G. Nair for designing database operations.

**Author Contributions** Work on the paper was divided between data production and analysis. The analysts were A.A., R.A., P.A., S.B., N.B., D.Z.C., C.C., D.C., Y.F., M.H., A.H., E.K., A.K., J.Le., R.M., X.J.M., J.R., A.S., J.W., K.-K.Y., K.Y.Y. and G.Z.-S. The data producers were N.A., A.P.B., P.C., A.C., Y.C., C.E., G.E., P.J.F., S.F., J.G., F.G., P.J., M.K., P.L., S.G.L., J.Li., H.M., R.M.M., H.O'G., Z.O., E.C.P., D.P., F.P., D.R., L.R., T.E.R., B.R., M.Sh., T.S., S.M.W., L.W. and X.Y. Larger efforts in analysis and data production are ascribed to the joint first authors. Author contributions to specific exhibits and files are shown in Supplementary Information, sections N and O. Overall project management was carried out by the two corresponding authors, M.B.G. and M.Sn.

**Author Information** Data sets described here can be obtained from the ENCODE project website at http://www.encodeproject.org and from http://encodenets.gersteinlab.org. More detail on data availability is in Supplementary Information, sections B and N. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and the online version of the paper is freely available to all readers. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.B.G. (mark.gerstein@yale.edu) or M.Sn. (mpsnyder@stanford.edu).

doi:10.1038/nature11233

# Landscape of transcription in human cells

Sarah Djebali[1]*, Carrie A. Davis[2]*, Angelika Merkel[1], Alex Dobin[2], Timo Lassmann[3], Ali Mortazavi[4,5], Andrea Tanzer[1],
Julien Lagarde[1], Wei Lin[2], Felix Schlesinger[2], Chenghai Xue[2], Georgi K. Marinov[4], Jainab Khatun[6], Brian A. Williams[4],
Chris Zaleski[2], Joel Rozowsky[7,8], Maik Röder[1], Felix Kokocinski[9], Rehab F. Abdelhamid[3], Tyler Alioto[1,10], Igor Antoshechkin[4],
Michael T. Baer[2], Nadav S. Bar[11], Philippe Batut[2], Kimberly Bell[2], Ian Bell[12], Sudipto Chakrabortty[2], Xian Chen[13],
Jacqueline Chrast[14], Joao Curado[1], Thomas Derrien[1], Jorg Drenkow[2], Erica Dumais[12], Jacqueline Dumais[12], Radha Duttagupta[12],
Emilie Falconnet[15], Meagan Fastuca[2], Kata Fejes-Toth[2], Pedro Ferreira[1], Sylvain Foissac[12], Melissa J. Fullwood[16], Hui Gao[12],
David Gonzalez[1], Assaf Gordon[2], Harsha Gunawardena[13], Cedric Howald[14], Sonali Jha[2], Rory Johnson[1], Philipp Kapranov[12,17],
Brandon King[4], Colin Kingswood[1,10], Oscar J. Luo[16], Eddie Park[5], Kimberly Persaud[2], Jonathan B. Preall[2], Paolo Ribeca[1,10],
Brian Risk[6], Daniel Robyr[15], Michael Sammeth[1,10], Lorian Schaffer[4], Lei-Hoon See[2], Atif Shahab[16], Jorgen Skancke[1,11],
Ana Maria Suzuki[3], Hazuki Takahashi[3], Hagen Tilgner[1]†, Diane Trout[4], Nathalie Walters[14], Huaien Wang[2], John Wrobel[6],
Yanbao Yu[13], Xiaoan Ruan[16], Yoshihide Hayashizaki[3], Jennifer Harrow[9], Mark Gerstein[7,8,18], Tim Hubbard[9],
Alexandre Reymond[14], Stylianos E. Antonarakis[15], Gregory Hannon[2], Morgan C. Giddings[6,13], Yijun Ruan[16], Barbara Wold[4],
Piero Carninci[3], Roderic Guigó[1,19] & Thomas R. Gingeras[2,12]

**Eukaryotic cells make many types of primary and processed RNAs that are found either in specific subcellular compartments or throughout the cells. A complete catalogue of these RNAs is not yet available and their characteristic subcellular localizations are also poorly understood. Because RNA represents the direct output of the genetic information encoded by genomes and a significant proportion of a cell's regulatory capabilities are focused on its synthesis, processing, transport, modification and translation, the generation of such a catalogue is crucial for understanding genome function. Here we report evidence that three-quarters of the human genome is capable of being transcribed, as well as observations about the range and levels of expression, localization, processing fates, regulatory regions and modifications of almost all currently annotated and thousands of previously unannotated RNAs. These observations, taken together, prompt a redefinition of the concept of a gene.**

**ENCODE**
Encyclopedia of DNA Elements
nature.com/encode

As the technologies for RNA profiling and for cell-type isolation and culture continue to improve, the catalogue of RNA types has grown and led to an increased appreciation for the numerous biological functions carried out by RNA, arguably putting them on par with the functional importance of proteins[1]. The Encyclopedia of DNA Elements (ENCODE) project has sought to catalogue the repertoire of RNAs produced by human cells as part of the intended goal of identifying and characterizing the functional elements present in the human genome sequence[2]. The five-year pilot phase of the ENCODE project[3] examined approximately 1% of the human genome and observed that the gene-rich and gene-poor regions were pervasively transcribed, confirming results of previous studies[4,5]. During the second phase of the ENCODE project, lasting 5 years, the scope of examination was broadened to interrogate the complete human genome. Thus, we have sought to both provide a genome-wide catalogue of human transcripts and to identify the subcellular localization for the RNAs produced. Here we report identification and characterization of annotated and novel RNAs that are enriched in either of the two major cellular subcompartments (nucleus and cytosol) for all 15 cell lines studied, and in three additional subnuclear compartments in one cell line. In addition, we have sought to determine whether identified transcripts are modified at their 5′ and 3′ termini by the presence of a 7-methyl guanosine cap or polyadenylation, respectively. We further studied primary transcript and processed product relationships for a large proportion of the previously annotated long and small RNAs. These results considerably extend the current genome-wide annotated catalogue of long polyadenylated and small RNAs collected by the GENCODE annotation group[6–8]. Taken together, our genome-wide compilation of subcellular localized and product-precursor-related RNAs serves as a public resource and reveals new and detailed facets of the RNA landscape.

- Cumulatively, we observed a total of 62.1% and 74.7% of the human genome to be covered by either processed or primary transcripts, respectively, with no cell line showing more than 56.7% of the union of the expressed transcriptomes across all cell lines. The consequent reduction in the length of 'intergenic regions' leads to a significant

---

[1] Centre for Genomic Regulation and UPF, Doctor Aiguader 88, Barcelona 08003, Catalonia, Spain. [2] Cold Spring Harbor Laboratory, Functional Genomics, 1 Bungtown Road, Cold Spring Harbor, New York 11742, USA. [3] RIKEN Yokohama Institute, RIKEN Omics Science Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. [4] California Institute of Technology, Division of Biology, 2 Beckman Institute, Pasadena, California 91125, USA. [5] University of California Irvine, Department of Developmental and Cell Biology, 2300 Biological Sciences III, Irvine, California 92697, USA. [6] Boise State University, College of Arts & Sciences, 1910 University Drive, Boise, Idaho 83725, USA. [7] Program in Computational Biology and Bioinformatics, Yale University, Bass 432, 266 Whitney Avenue, New Haven, Connecticut 06520, USA. [8] Department of Molecular Biophysics and Biochemistry, Yale University, Bass 432, 266 Whitney Avenue, New Haven, Connecticut 06520, USA. [9] Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. [10] Centro Nacional de Análisis Genómico (CNAG), C/ Baldiri Reixac 4, Torre I, Barcelona 08028, Catalonia, Spain. [11] Department of Chemical Engineering, Norwegian University of Science and Technology, Trondheim NO-7491, Norway. [12] Affymetrix, Inc, 3380 Central Expressway, Santa Clara, California 95051, USA. [13] University of North Carolina at Chapel Hill, Department of Biochemistry & Biophysics, 120 Mason Farm Road, Chapel Hill, North Carolina 27599, USA. [14] University of Lausanne, Center for Integrative Genomics, Genopode building, Lausanne 1015, Switzerland. [15] University of Geneva Medical School, Department of Genetic Medicine and Development and iGE3 Institute of Genetics and Genomics of Geneva, 1 rue Michel-Servet, Geneva 1211, Switzerland. [16] Genome Institute of Singapore, Genome Technology and Biology, 60 Biopolis Street, 02-01, Genome, Singapore 138672, Singapore. [17] St Laurent Institute, One Kendall Square, Cambridge, Massachusetts 02141, USA. [18] Department of Computer Science, Yale University, Bass 432, 266 Whitney Avenue, New Haven, Connecticut 06520, USA. [19] Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Catalonia, Spain. †Present address: Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA.
*These authors contributed equally to this work.

overlapping of neighbouring gene regions and prompts a redefinition of a gene.

- Isoform expression by a gene does not follow a minimalistic expression strategy, resulting in a tendency for genes to express many isoforms simultaneously, with a plateau at about 10–12 expressed isoforms per gene per cell line.
- Cell-type-specific enhancers are promoters that are differentiable from other regulatory regions by the presence of novel RNA transcripts, chromatin marks and DNase I hypersensitive sites.
- Coding and non-coding transcripts are predominantly localized in the cytosol and nucleus, respectively, with a range of expression spanning six orders of magnitude for polyadenylated RNAs, and five orders of magnitude for non-polyadenylated RNAs.
- Approximately 6% of all annotated coding and non-coding transcripts overlap with small RNAs and are probably precursors to these small RNAs. The subcellular localization of both annotated and unannotated short RNAs is highly specific.

## RNA data set generation

We performed subcellular compartment fractionation (whole cell, nucleus and cytosol) before RNA isolation in 15 cell lines (Supplementary Table 1) to interrogate deeply the human transcriptome. For the K562 cell line, we also performed additional nuclear subfractionation into chromatin, nucleoplasm and nucleoli. The RNAs from each of these subcompartments were prepared in replica and were separated based on length into >200 nucleotides (long) and <200 nucleotides (short). Long RNAs were further fractionated into polyadenylated and non-polyadenylated transcripts. A number of complementary technologies were used to characterize these RNA fractions as to their sequence (RNA-seq), sites of initiation of transcription (cap-analysis of gene expression (CAGE)[9]) and sites of 5′ and 3′ transcript termini (paired end tags (PET)[10]; Supplementary Fig. 1). Sequence reads were mapped and post-processed using a variety of software tools (Supplementary Table 2 and Supplementary Fig. 2). We used the mapped data to assemble and quantify *de novo* elements (exons, transcripts, genes, contigs, splice junctions and transcription start sites (TSSs)) as well as to quantify annotated GENCODE (v7) elements. Elements and quantifications were further assessed for reproducibility between replicates using a non-parametric version (npIDR, Supplementary Information) of the irreproducible detection rate (IDR) statistical test[11]. Only elements deemed to be reproducible with at least 90% likelihood were used in most analyses. The raw data, mapped data and elements were then made available by the ENCODE Data Coordination Center (DCC, http://genome.ucsc.edu/ENCODE/dataSummary.html) (Supplementary Fig. 2). These data, as well as additional data on all intermediate processing steps, are available on the RNA Dashboard (http://genome.crg.cat/encode_RNA_dashboard/).

## Long RNA expression landscape

### Detection of annotated and novel transcripts

The GENCODE gene (Supplementary Fig. 3a) and transcript (Supplementary Fig. 3b) reference annotation[8] captures our current understanding of the polyadenylated human transcriptome. In the samples interrogated here, we cumulatively detected 70% of annotated splice junctions, transcripts and genes (Fig. 1 and Table 1a). We also detected approximately 85% of annotated exons with an average coverage by RNA-seq contigs of 96%. The variation in the proportion of detected elements among cell lines was small (Fig. 1, width of box plots). Consistent with earlier studies, most annotated elements are present in both polyadenylated (Supplementary Table 3a) and non-polyadenylated (Supplementary Table 3b) samples[12–15]. Only a small proportion of GENCODE elements (0.4% of exons, 2.8% of splice sites, 3.3% of transcripts and 4.7% of genes) are detected exclusively in the non-polyadenylated RNA fraction.

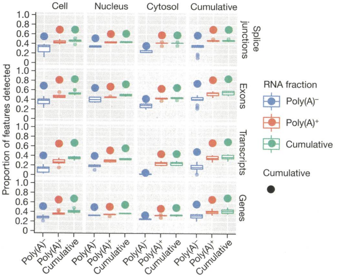


**Figure 1 | A large majority of GENCODE elements are detected by RNA-seq data.** Shown are GENCODE-detected elements in the polyadenylated and non-polyadenylated fractions of cellular compartments (cumulative counts for both RNA fractions and compartments refer to elements present in any of the fractions or compartments). Each box plot is generated from values across all cell lines, thus capturing the dispersion across cell lines. The largest point shows the cumulative value over all cell lines.

Beyond the GENCODE annotated elements, we observed a substantial number of novel elements represented by reproducible RNA-seq contigs. These novel elements covered 78% of the intronic nucleotides and 34% of the intergenic sequences (Supplementary Fig. 4). Overall, the unique contribution of each cell line to the coverage of the genome tends to be small and similar for each cell line (Supplementary Fig. 5). We used the Cufflinks algorithm (see Supplementary Information), and predicted over all long RNA-seq samples 94,800 exons, 69,052 splice junctions, 73,325 transcripts and 41,204 genes in intergenic and antisense regions (Table 1b). These novel elements increase the GENCODE collection of exons, splice sites, transcripts and genes by 19%, 22%, 45% and 80%, respectively. The increase in the number of genes and the relatively low contribution of novel splice sites is primarily caused by the detection of both polyadenylated and non-polyadenylated mono-exonic transcripts (Supplementary Table 3). Detection of unspliced transcripts could partially be an artefact caused by low levels of DNA contamination or by incomplete determination of transcript structures.

Independent validation of multi-exonic transcript models and the associated predicted coding products were carried out using overlapping targeted 454 Life Sciences (Roche) paired-end reads and mass spectrometry. Of approximately 3,000 intergenic and antisense transcript models tested, validation rates from 70% to 90% were observed, depending on the number of reads and IDR score. In addition, these experiments led to the identification of more than 22,000 novel splice sites not previously detected, meaning an almost eightfold increase in detection compared to the sites originally detected with RNA-seq (Supplementary Fig. 6). Using mass spectrometric analyses, we investigated what fraction of the novel Cufflinks transcript models show evidence consistent with protein expression. We produced 998,570 spectra from two cell lines (K562 and GM12878; J. Khatun *et al.*, manuscript in preparation), and mapped them to a three-frame translation of the novel Cufflinks models (Supplementary Material). At a 1% false discovery rate (FDR), we identified 419 novel models with 5 or more spectral and/or 2 or more peptide hits, of which only 56 were intergenic or antisense to GENCODE genes (Supplementary Table 4 and Supplementary Fig. 7). Thus, most novel transcripts seem to lack protein-coding capacity.

### The transcriptome of nuclear subcompartments

For the K562 cell line, we also analysed RNA isolated from three subnuclear compartments (chromatin, nucleolus and nucleoplasm;

Supplementary 5). Almost half (18,330) of the GENCODE (v7) annotated genes detected for all 15 cell lines (35,494) were identified in the analysis of just these three nuclear subcompartments. In addition, there were as many novel unannotated genes found in K562 subcompartments as there were in all other data sets combined (Supplementary Table 5 and Table 1b). For all annotated (Supplementary Table 5.1) or novel (Supplementary Table 5.2) elements, only a small fraction in each subcompartment was unique to that compartment (Supplementary Table 6).

The interrogation of different subcellular RNA fractions provides snapshots of the status of the RNA population along the RNA processing pathway. Thus, by analysing short and long RNAs in the different subcellular compartments, we confirm that splicing predominantly occurs during transcription. By using RNA-seq to measure the degree of completion of splicing (Fig. 2a), we observed that around most exons, introns are already being spliced in chromatin-associated RNA—the fraction that includes RNAs in the process of being transcribed (Fig. 2b). Concomitantly, we found strong enrichment specifically of spliceosomal small nuclear RNAs (snRNAs) in this RNA fraction (see 'Short RNA expression landscape' later). Co-transcriptional splicing provides an explanation for the increasing evidence connecting chromatin structure to splicing regulation, and we have observed that exons in the process of being spliced are enriched in a number of chromatin marks[16,17].

## Gene expression across cell lines

The analyses of RNAs isolated from different subcellular compartments also provide information concerning compartment-specific relative steady-state abundance and the post transcriptional processing state (spliced/unspliced, polyadenylated/non-polyadenylated, 5′ capped/uncapped) for each of the detected transcripts. The observed range of gene expression spans six orders of magnitude for polyadenylated RNAs (from $10^{-2}$ to $10^{4}$ reads per kilobase per million reads (r.p.k.m.)), and five orders of magnitude (from $10^{-2}$ to $10^{3}$ r.p.k.m.) for non-polyadenylated RNAs (Fig. 3 and Supplementary Fig. 8a). The distribution of gene expression is very similar across cell lines, with protein-coding genes, as a class, having on average higher expression levels than long non-coding RNAs (lncRNAs). Assuming that 1–4 r.p.k.m. approximates to 1 copy per cell[18], we find that almost one-quarter of expressed protein-coding genes and 80% of the detected lncRNAs are present in our samples in 1 or fewer copies per cell. The general lower level of gene expression measured in lncRNAs may not necessarily be the result of consistent low RNA copy number in all cells within the population interrogated, but may also result from restricted expression in only a subpopulation of cells. In some cell lines, individual lncRNAs can exhibit steady-state expression levels as high as those of protein-coding genes. This is, for example, seen in the expression of the protein-coding gene actin, gamma 1 (*ACTG1*), and the non-coding gene, *H19* (Fig. 3). *ACTG1* transcripts are part of all non-muscle cytoskeleton systems within cells and show a steady-state expression level at the population level that is at least 1–2 logs greater than *H19*, a cytosolic non-coding RNA (ncRNA). However, when measured at the individual transcript level, expression of lncRNA transcripts is comparable to that of individual protein-coding transcripts (Supplementary Fig. 8b).

Novel antisense and intergenic genes predicted in this study comprise a third clustering of RNAs with levels of expression ranging from $10^{-4}$ to $10^{-1}$ r.p.k.m. As a class, only protein-coding genes seem to be enriched in the cytosol, making the nucleus a centre for the accumulation of ncRNAs (Fig. 3). Other gene classes, such as pseudogenes and small annotated ncRNAs, also show subcellular compartmental enrichment (Supplementary Fig. 9).

Higher variability and lower pairwise correlation of expression across all cell lines is consistent with lncRNAs contributing more to cell-line specificity than protein-coding genes. Indeed, a considerable fraction (29%) of all expressed lncRNAs are detected in only one of the cell lines studied when considering the whole cell polyadenylated RNAs, whereas only 10% were expressed in all cell lines. Conversely, whereas a large fraction (53%) of expressed protein-coding genes were constitutive (expressed in all cell lines), only ~7% were cell-line specific (Supplementary Table 7 and Supplementary Fig. 10).

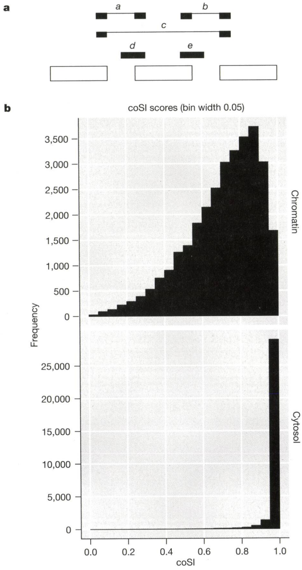


**Figure 2 | Co-transcriptional splicing. a**, Short read mappings for exon-based splicing completion. Read mappings that allow assessment of splicing completion around exons are shown. Reads providing evidence of splicing completion for the region containing the exon (with either exon inclusion ($a$, $b$) or exclusion ($c$)) are shown. Reads providing evidence for the splicing of the region containing the exon not being completed yet are indicated by $d$ and $e$. The complete splicing index (coSI) is the ratio of $(0.5(a + b) + c)$ over $(0.5(a + b) + c + 0.5(d + e))$ and can thus be broadly assumed to correspond to the fraction of RNA molecules in which the region containing the exon has already been spliced (see ref. 16). A coSI value of 1 means splicing completed, whereas a value of 0 indicates that splicing has not yet been initiated. **b**, Distribution of coSI scores computed on GENCODE internal exons. Top: distribution in the total chromatin RNA fraction. Bottom: distribution in cytosolic polyadenylated RNA fraction.

## Patterns of splicing

The analysis of the expression of alternative isoforms resulted in several observations. First, isoform expression does not seem to follow a minimalistic strategy. Genes tend to express many isoforms simultaneously, and as the number of annotated isoforms per gene grows, so does the number of expressed isoforms (Fig. 4a). The increase, however, is not linear and seems to plateau at about 10–12 expressed

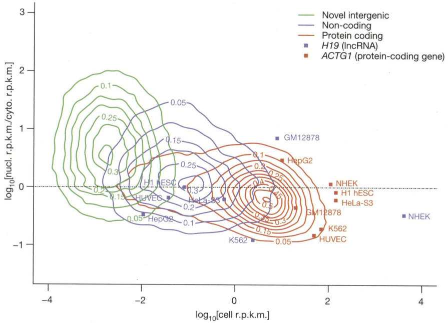


**Figure 3 | Abundance of gene types in cellular compartments.** Two-dimensional kernel density plots of nuclear over cytosolic enrichment (*y* axis) versus overall gene expression in the whole cell extract (*x* axis), for protein coding, long non-coding and novel genes over all cell lines. Only genes present in all three RNA extracts are displayed, as well as two representative genes (*ACTG1* in red and *H19* in blue), for which the expression in each individual cell line is shown. The actual values of the estimated kernel density are indicated by contour lines and colour shades.

isoforms per gene. However, we cannot obviously distinguish whether this is the result of multiple isoforms expressed in the same cell or of different isoforms expressed in different cells within the interrogated population. Second, alternative isoforms within a gene are not expressed at similar levels, and one isoform dominates in a given condition—usually capturing a large fraction of the total gene expression (at least 30%, even for genes with many isoforms; Fig. 4b). Third, about three-quarters of protein-coding genes have at least two different dominant/major isoforms depending on the cell line (Supplementary Fig. 11a). Fourth, the number of major isoforms per gene grows with the number of annotated isoforms; indeed, the proportion of genes with *n* isoforms that express only one major isoform is strikingly proportional to 1/*n* (Supplementary Fig. 11b). Fifth, variability of gene expression contributes more than variability of splicing ratios to the variability of transcript abundances across cell lines (Supplementary Information).

## Alternative transcription initiation and termination

On the basis of RNA-seq analysis of polyadenylated RNAs, a total of 128,021 TSSs were detected across all cell lines, of which 97,778 were previously annotated and 30,243 were novel intergenic/antisense TSSs (Supplementary Table 3a). CAGE tags, filtered by a hidden Markov model (HMM)-based algorithm to differentiate between 5′ capped termini of polymerase II transcripts and recapping events[19] (Supplementary Information), identified a total of 82,783 non-redundant TSSs (Supplementary Table 8). Approximately 48% of the CAGE-identified TSSs are located within 500 base pairs (bp) of an annotated RNA-seq-detected GENCODE TSS, whereas an additional 3% are within 500 bp of a novel TSS (Supplementary Fig. 12). Notably, only ~72% of all CAGE sequencing reads map to TSSs, indicating that the remaining 30% may originate from recapping events or from a new class of TSS.

Using data collected within the ENCODE consortium[20], we carried out a comparison of the GENCODE/RNA-seq and CAGE-determined TSSs and correlated them to chromatin and DNA features characteristic of initiation of transcription, such as DNase hypersensitivity[21], chromatin modification and DNA binding elements[22,23]. All GENCODE/RNA-seq-determined TSSs were examined in each of the cell lines (Supplementary Fig. 13, column 1). Of these redundant positions, 44.7% (199,146) of the RNA-seq-supported TSSs also displayed

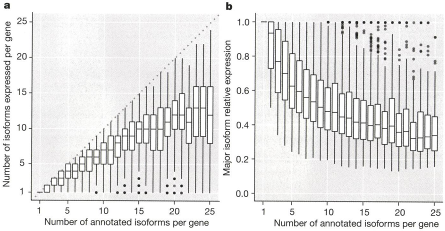


**Figure 4 | Isoform expression within a gene.** **a**, Number of expressed isoforms per gene per cell line. Genes tend to express many isoforms simultaneously. **b**, Relative expression of the most abundant isoform per gene per cell line. There is generally one dominant isoform in a given condition. The whiskers are defined as Q1 −1.5 × IQR to Q3 +1.5 × IQR, where IQR is the interquartile range, and Q1 and Q3 the first and third quartile, respectively. Each box plot was constructed using the number of genes with 1, 2, 3, 4, etc. up to 25 isoforms.

evidence of CAGE. Approximately half of these TSS positions are associated with at least one of the other characteristic features of transcription initiation (DNase I, H3K27ac and H3K4me3 chromatin modifications). Thus, only a small minority of the TSSs identified by either CAGE or RNA-seq/GENCODE displayed all of the characteristics of the start of transcription (presence of DNase I, H3K4me3, H3K27ac sites and either TAF1 or TBP binding). This is consistent with the possibility that regulatory regions proximal to TSSs are of more than one type.

At the 3′ end, a total of 128,824 sites mapping within annotated GENCODE transcripts were identified as potential sites of polyadenylation after trimming unmapped RNA-seq reads with long terminal polyadenine stretches[24]. About 20% of these mapped proximal to annotated polyadenylation sites (PAS) whereas the remaining 80% correspond to novel PAS of annotated genes, raising the average number of PAS per gene from 1.1 to 2.5. Generally, we observed a cell-type preference for proximal PAS (closest to the annotated stop codon) in the cytosol compared to the nucleus (Supplementary Information).

## Short RNA expression landscape

### Annotated small RNAs

Currently, a total of 7,053 small RNAs are annotated by GENCODE, 85% of which correspond to four major classes: small nuclear (sn)RNAs, small nucleolar (sno)RNAs, micro (mi)RNAs and transfer (t)RNAs (Table 2a). Overall we find 28% of all annotated small RNAs to be expressed in at least one cell line (Table 2a). The distribution of annotated small RNAs differs markedly between cytosolic and nuclear compartments (Supplementary Fig. 14a). We found that the small RNA classes were enriched in those compartments where they are known to perform their functions: miRNAs and tRNAs in the cytosol, and snoRNAs in the nucleus. Interestingly, snRNAs were equally abundant in both the nucleus and the cytosol. When specifically interrogating the subnuclear compartments of the K562 cell line, however, snRNAs seem to be present in very high abundance in the chromatin-associated RNA fraction (Supplementary Fig. 14b, c). This striking enrichment is consistent with splicing being predominantly co-transcriptional[16,25].

### Unannotated short RNAs

We detected two types of unannotated short RNAs. The first type corresponds to subfragments of annotated small RNAs. Because we performed 36-nucleotide end-sequencing of the small RNA fraction, we expected RNA-seq reads to map to the 5′ end of the small RNAs. Supplementary Figure 15 shows the mapping profile of reads along small RNA genes. In both the nuclear and cytosolic compartments, we indeed detected accumulation of reads at the start of snoRNAs and at the guide and passenger sequences of annotated miRNAs. For snRNAs, however, we observed three prominent peaks: the expected one at the 5′ end and two smaller ones at the middle and at the 3′ end of the gene, indicating fragmentation of some snRNAs. Finally, tRNAs seem not to have any prominent sets of 5′ end fragments present at levels greater than what is seen at the annotated 5′ termini. Whereas subfragments of mature tRNAs have been reported previously, these reports were confined to distinct alleles of only a few tRNA genes[26–28].

The second and largest source of unannotated short RNAs corresponds to novel short RNAs (Table 2b) that map outside of annotated ones. Almost 90% of these are only observed in one cell line and are present at low copy numbers. Nearly 40% of these unannotated short RNAs are associated with promoter and terminator regions of annotated genes (promoter-associated short RNAs (PASRs) and termini-associated short RNAs (TASRs)), and their position relative to TSSs and transcription termination sites is similar to previous results[4].

### Genealogy of short RNAs

Genome wide, 27% of annotated small RNAs reside within 8% of protein-coding and 5% within 3% of lncRNA genes (Supplementary Fig. 16). Overall, about 6% of all annotated long transcripts overlap with small RNAs and are probably precursors to these small RNAs. Although most of these small RNAs reside in introns, when controlling for relative exon/intron length, we found that exons from lncRNAs are comparatively enriched as hosts for snoRNAs (Supplementary Fig. 17a). Additionally, 8.4% of GENCODE annotated small RNAs map within novel intergenic transcripts, with most overlapping annotated tRNAs. The enrichment for tRNAs was mostly in novel intergenic transcripts derived from non-polyadenylated RNAs (Supplementary Fig. 17b). Many long RNAs, both novel and annotated, thus seem to have dual roles, as functional (protein coding) RNAs, and as precursors for many important classes of small RNAs. Using RNA-seq data from the K562 cell line, we investigated the preferential cellular localization of these RNA precursors (Supplementary Fig. 18). For mature miRNAs and tRNAs (cytosolic enrichment), the potential RNA precursors, identified as RNA-seq contigs overlapping the small RNAs, were detected to be predominantly nuclear (Supplementary Fig. 18a, d). Notably, whereas mature snRNAs were both nuclear and cytosolic, the overlapping long RNAs were observed to be primarily nuclear (Supplementary Fig. 18c). Finally, for snoRNAs (nuclear enrichment), potential long RNA precursors were decidedly observed to be both nuclear and cytosolic (Supplementary Fig. 18b). Unannotated short RNAs were found overall not to be enriched in either the nuclear or cytosolic compartment (Supplementary Fig. 18e).

## RNA editing and allele-specific expression

The sequence of transcripts can differ from the underlying genomic sequence as the result of post-transcriptional editing. We developed a pipeline to filter sequencing artefacts and identify genes that are RNA edited[29]. Focusing first on GM12878, a cell line that has been deeply re-sequenced, we find a total 51,557 RNA consistent single nucleotide variants (SNVs) within genic boundaries, 65% of which are present in dbSNP. Of the remainder, 1,186 SNVs in 430 genes (Supplementary Fig. 19a) survive our most stringent filters and 88% of these are candidate adenosine to inosine A>G(I) changes. Notably, the next highest frequency of SNVs is for T>C (5%) and these occur primarily in regions with detectable antisense transcription[29]. We find similar A>G(I) frequencies of 75–84% in seven additional cell lines (Supplementary Fig. 19b). The remaining non-canonical edits amount to very few events in each cell line and are relatively evenly distributed (G>A is the third highest). These results do not support a recent report of a substantial number of non-canonical SNV edits in the RNA of human lymphoblastoid cells[30].

Using the AlleleSeq pipeline[31] on the SNPs in the GM12878 genome, we found that approximately 18% of both GENCODE annotated protein-coding and long non-coding genes exhibit allele-specific expression. The proportion of genes with allele-specific expression was similar in the three investigated RNA fractions (whole-cell, cytoplasm and nucleus; Supplementary Table 9 and Supplementary Information).

## Repeat region transcription

About 18% (14,828) of CAGE-defined TSS regions overlap repetitive elements. More precisely, we find 322, 315, 507 and 1,262 intergenic CAGE clusters overlapping long interspersed element (LINE), short interspersed element (SINE), long terminal repeat (LTR) and other repeat elements, respectively (see Supplementary Information). Measuring Shannon entropy across cell lines, we found that CAGE clusters mapping to repeat regions were noticeably more narrowly expressed than CAGE clusters mapping within genic regions (Supplementary Fig. 20a). We represented the correlation of levels of expression compared to cell types as heat maps drawn separately for each of the three repeat element families (LINE, SINE and LTR) (Supplementary Fig. 20b–d). Although a large proportion of the transcripts in the human genome is thought to be initiated from repetitive elements (especially retrotransposon elements[32]), these data clearly point to cell-line specificity as the main characteristic of transcripts emanating from repeat regions.

## Characterization of enhancer RNA

It has recently been reported that RNA polymerase II binds some distal enhancer regions and can produce enhancer-associated transcripts named eRNA[33–35]. We used our RNA assays to detect and characterize transcriptional activity at enhancer loci predicted genome-wide from ENCODE chromatin immunoprecipitation and high-throughput sequencing (ChIP-seq) data[20,36].

Figure 5a shows the aggregate pattern of RNA-seq and CAGE signal in a strand-specific manner around the subset of predicted gene-distal enhancers containing DNase I hypersensitive sites and centred on those sites. In these plots, as denoted by the accumulation of CAGE tags signifying TSSs, transcription initiation within the enhancer region is observed, and continues outwards for several kilobases (kb). This behaviour can be observed for the polyadenylated and non-polyadenylated RNA fractions mapping in both intronic and intergenic regions. As previously reported[33], we observe a large diversity of expression levels at each of the transcribed enhancers. Polyadenylated to non-polyadenylated RNA ratios, as well as nuclear to cytoplasmic ratios, vary at individual enhancers (Supplementary Fig. 21a, b). However, contrary to some previous reports, although most eRNAs are prevalent in the nuclear non-polyadenylated RNA fraction, some eRNAs seemed to be polyadenylated in the nucleus. This pattern was significantly different compared to transcripts from GENCODE annotated and novel predicted[20] promoters (Fig. 5b).

Transcribed enhancers on average show a significantly different pattern of chromatin modification than non-transcribed ones[37–40]. The enhancer regions displayed stronger signals for H3K4 methylation, H3K27 acetylation and H3K79 dimethylation along with higher levels of RNA polymerase II binding, all associated with

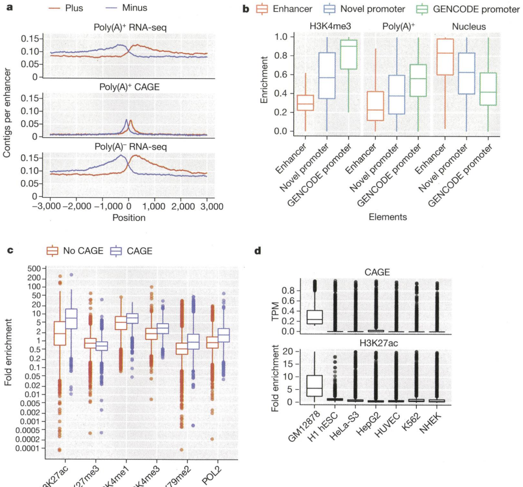


**Figure 5 | Transcription at enhancers. a**, The pattern of RNA elements around enhancer predictions[20,36] containing DNase I hypersensitive sites. The lines represent the average frequency of RNA elements (top, polyadenylated long RNA contigs; middle, CAGE tag clusters; bottom, non-polyadenylated long RNA contigs) in a genomic window around the centre of the enhancer prediction as determined by DNase I hypersensitive sites. Elements on the plus strand are shown in red, and on the minus strand in blue. **b**, Enhancer transcripts differ from promoter transcripts. The box plots compare the features of transcripts at predicted enhancer loci compared to predicted novel intergenic promoters[20] and annotated promoters[8]. H3K4me3, poly(A)$^+$ and nucleus denote the three following ratios: H3K4me3/(H3K4me3 + H3K4me1), polyadenylated/(polyadenylated + non-polyadenylated), nuclear/(nuclear + cytosolic). Enhancers are marked by higher levels of H3K4me1 compared to H3K4me3 than novel or annotated promoters (left). Enhancer transcripts show higher levels of non-polyadenylated (middle) and nuclear (right) RNA relative to promoters. **c**, Chromatin state at transcribed enhancers. Enhancer predictions with evidence of transcription (in blue; Cage tags present at predicted locus) show a different pattern of histone modification and higher levels of RNA polymerase II binding than non-transcribed predictions (red). They are enriched for H3K27 acetylation, H3K4 methylation, H3K79 dimethylation and depleted for H3K27 trimethylation. **d**, Enhancer activity and transcription is cell-type specific. Loci predicted to be active transcribed enhancers in GM12878 cells show low signal for CAGE tags (top) and for H3K27 acetylation (bottom) in other cell lines. The whiskers are defined as Q1 −1.5 × IQR to Q3 +1.5 × IQR, where IQR is the interquartile range, and Q1 and Q3 the first and third quartile, respectively.

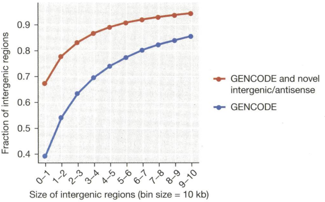


**Figure 6 | Size distribution of intergenic regions.** Novel genes increase the proportion of small intergenic regions.

transcriptional initiation and elongation (Fig. 5c). Both the transcripts and the chromatin states are cell-type specific (Fig. 5d). Taking the GM12878 cell line as an example, the enhancer loci producing eRNA demonstrate enrichment of CAGE tag detection (Fig. 5d, top) and the presence of H3K27ac histone modification (Fig. 5d, bottom) in this cell line compared to five other analysed cell lines. This strongly suggests that the regulatory regions governing the expression of enhancer transcripts are distinguished from regulatory regions located at the beginning of genic regions.

## Concluding remarks

The cumulative coverage of transcribed regions in the 15 cell lines across the human genome is 62.1% and 74.7% for processed and primary transcripts, respectively (Supplementary Table 10 and Supplementary Fig. 22). On average, for each cell line, 39% of the genome is covered by primary transcripts and 22% by processed RNAs. No cell line showed transcription of more than 56.7% of the union of the expressed transcriptomes across all cell lines. When mapping the current RNA-seq data to the ENCODE pilot regions (Supplementary Table 10), we observed a similar, albeit higher, extent of transcriptional coverage of 73.3% for processed RNAs and 84.5% for primary transcripts. Previously reported estimates in these regions for processed and primary transcripts were 24% and 93%, respectively (Supplementary Table 2.4.3 and ref. 3). The increased genome coverage by processed RNAs stems largely from the inclusion of

**Table 1 | Long polyadenylated and non-polyadenylated RNAs**

Expression of GENCODE (v7) annotated elements (a)

| Gene type | Detected exons† (annotation no.) | Detected splice junctions† (annotation no.) | Detected transcripts† (annotation no.) | Detected genes† (annotation no.) | Exon nucleotide coverage‡ (%) | Number of genes expressed in at least one cell line | Number of genes expressed in only one cell line | Proportion over genes expressed§ (%) | Number of genes expressed in 14 cell lines | Proportion over genes expressed‖ (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| Long non-coding | 22,381 (41,467) | 8,017 (26,872) | 6,521 (14,880) | 5,906 (9,277) | 87.5 | 5,906 | 1,386 | 23.5 | 631 | 10.7 |
| Protein coding | 288,322 (318,514) | 194,752 (244,158) | 59,822 (76,006) | 18,939 (20,679) | 98.1 | 18,939 | 1,082 | 5.7 | 10,571 | 55.8 |
| Other* | 102,000 (133,937) | 19,277 (47,663) | 45,410 (71,113) | 10,649 (21,750) | 95.2 | 10,649 | 2,453 | 23.0 | 1,896 | 17.8 |
| Total annotated | 412,703 (493,918) | 222,046 (318,693) | 111,753 (161,999) | 35,494 (51,706) | 96.7 | 35,394 | 4,921 | 13.9 | 13,098 | 37.0 |

Expression of GENCODE (v7) intergenic and antisense elements (b)

| Category | Detected exons† | Detected splice junction† | Detected transcripts† | Detected genes† |
|---|---|---|---|---|
| Mono-exonic | 55,683 | NA | 55,682 | 33,686 |
| Multi-exonic | 39,117 | 69,052 | 17,643 | 7,518 |
| Total | 94,800 | 69,052 | 73,325 | 41,204 |

NA, not applicable.
* Includes pseudogenes, miRNAs, etc.
† All elements that passed npIDR (0.1).
‡ Cumulative detected nucleotide in detected exons/total nucleotides in detected exons.
§ Proportion for genes expressed in only one cell line.
‖ Proportion for genes expressed in 14 cell lines.

**Table 2 | Short RNAs**

Expression of GENCODE (v7) annotated small RNA genes (a)

| Gene type* | GENCODE total | Detected genes (% detected) | No. genes expressed in only one cell line (% detected) | No. genes expressed in 12 cell lines (% detected) | miRNA guide fragment‡ | miRNA passenger fragment§ | Internal fragments‖ of annotated small RNA (average per detected gene) |
|---|---|---|---|---|---|---|---|
| miRNA | 1,756 | 497 (28) | 59 (12) | 147 (30) | 454 (454) | 175 (175) | 18 |
| snoRNA | 1,521 | 458 (30) | 73 (16) | 223 (49) | NA | NA | 60 |
| snRNA | 1,944 | 378 (19) | 123 (33) | 41 (11) | NA | NA | 36 |
| tRNA | 624 | 465 (75) | 29 (6) | 197 (42) | NA | NA | 52 |
| Other† | 1,209 | 191 (16) | 69 (36) | 24 (13) | NA | NA | 32 |
| Total GENCODE | 7,054 | 1,989 (28) | 353 (18) | 632 (32) | NA | NA | 40 |

Expression of unannotated short RNAs (b)

| Cell compartment | Unannotated short RNAs | Exonic | Intronic | Exon–intron boundaries | Genic | Gene–intergene boundaries | Intergenic |
|---|---|---|---|---|---|---|---|
| Cell | 57,393 | 14,116 | 13,773 | 1,818 | 29,707 | 13,048 | 25,906 |
| Nucleus | 82,297 | 19,334 | 40,136 | 5,248 | 64,718 | 7,417 | 16,289 |
| Cytosol | 25,455 | 6,183 | 5,605 | 665 | 12,453 | 6,631 | 12,447 |
| Three compartments | 150,165 | 38,969 | 55,061 | 7,552 | 101,582 | 23,185 | 45,081 |

NA, not applicable.
* Includes all other GENCODE small transcript biotypes except for pseudogenes.
† All elements that have passed npIDR (0.1).
‡ Number of detected miRNAs with an expressed annotated guide (with an annotated guide in mirbase).
§ Number of detected miRNAs with an expressed annotated passenger (with an annotated passenger in mirbase).
‖ Short RNA-seq mapping for which the 5′ end starts 5 bp after the start and ends 5 bp before the end of a detected gene.

non-polyadenylated RNAs in the current study. Other than that, given the differences in the samples studied, the selection of pilot regions with high genic content, the increase of annotated genomic regions over time, and the different technologies used to interrogate transcription, both estimates are in reasonable agreement.

As a consequence of both the expansion of genic regions by the discovery of new isoforms and the identification of novel intergenic transcripts, there has been a marked increase in the number of intergenic regions (from 32,481 to 60,250) due to their fragmentation and a decrease in their lengths (from 14,170 bp to 3,949 bp median length; Fig. 6). Concordantly, we observed an increased overlap of genic regions. As the determination of genic regions is currently defined by the cumulative lengths of the isoforms and their genetic association to phenotypic characteristics, the likely continued reduction in the lengths of intergenic regions will steadily lead to the overlap of most genes previously assumed to be distinct genetic loci. This supports and is consistent with earlier observations of a highly interleaved transcribed genome[12], but more importantly, prompts the reconsideration of the definition of a gene. As this is a consistent characteristic of annotated genomes, we would propose that the transcript be considered as the basic atomic unit of inheritance. Concomitantly, the term gene would then denote a higher-order concept intended to capture all those transcripts (eventually divorced from their genomic locations) that contribute to a given phenotypic trait. Co-published ENCODE-related papers can be explored online via the *Nature* ENCODE explorer (http://www.nature.com/ENCODE), a specially designed visualization tool that allows users to access the linked papers and investigate topics that are discussed in multiple papers via thematically organized threads.

## METHODS SUMMARY

For full details of Methods, see Supplementary Information.

**Received 10 December 2011; accepted 15 May 2012.**

1. Mattick, J. S. Long noncoding RNAs in cell and developmental biology. *Semin. Cell Dev. Biol.* **22,** 327 (2011).
2. The ENCODE Project Consortium.. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306,** 636–640 (2004).
3. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447,** 799–816 (2007).
4. Kapranov, P. *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316,** 1484–1488 (2007).
5. Kapranov, P., Willingham, A. T. & Gingeras, T. R. Genome-wide transcription and the implications for genomic organization. *Nature Rev. Genet.* **8,** 413–423 (2007).
6. Coffey, A. J. *et al.* The GENCODE exome: sequencing the complete human exome. *Eur. J. Hum. Genet.* **19,** 827–831 (2011).
7. Harrow, J. *et al.* GENCODE: producing a reference annotation for ENCODE. *Genome Biol.* **7** (suppl. 1), 1–9 (2006).
8. Harrow, J. *et al.* GENCODE: The reference human genome annotation for the ENCODE project. *Genome Res.* (in the press).
9. Kodzius, R. *et al.* CAGE: cap analysis of gene expression. *Nature Methods* **3,** 211–222 (2006).
10. Ng, P. *et al.* Gene identification signature (GIS) analysis for transcriptome characterization and genome annotation. *Nature Methods* **2,** 105–111 (2005).
11. Li, Q., Brown, J. B., Huang, H. & Bickel, P. J. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5,** 1752–1779 (2011).
12. Cheng, J. *et al.* Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308,** 1149–1154 (2005).
13. Katinakis, P. K., Slater, A. & Burdon, R. H. Non-polyadenylated mRNAs from eukaryotes. *FEBS Lett.* **116,** 1–7 (1980).
14. Milcarek, C., Price, R. & Penman, S. The metabolism of a poly(A) minus mRNA fraction in HeLa cells. *Cell* **3,** 1–10 (1974).
15. Salditt-Georgieff, M., Harpold, M. M., Wilson, M. C. & Darnell, J. E. Jr. Large heterogeneous nuclear ribonucleic acid has three times as many 5′ caps as polyadenylic acid segments, and most caps do not enter polyribosomes. *Mol. Cell. Biol.* **1,** 179–187 (1981).
16. Tilgner, H. *et al.* Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* (in the press).
17. Tilgner, H. *et al.* Genomic analysis of ENCODE data reveals widespread links between epigenetic chromatin marks and alternative splicing. *Genome Res.* (in the press).
18. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5,** 621–628 (2008).
19. Affymetrix/Cold Spring Harbor Laboratory ENCODE Transcriptome Project. Post-transcriptional processing generates a diversity of 5′-modified long and short RNAs. *Nature* **457,** 1028–1032 (2009).
20. ENCODE Project Consortium.. An integrated encyclopaedia of DNA elements in the human genome. *Nature* http://dx.doi.org/10.1038/nature11247 (this issue).
21. Thurman, R. E. The accessible chromatin landscape of the human genome. *Nature* http://dx.doi.org/10.1038/nature11232 (this issue).
22. Gerstein, M. B. Architecture of the human regulatory network derived from ENCODE data. *Nature* http://dx.doi.org/10.1038/nature11245 (this issue).
23. Wang, J. *et al.* Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.* (in the press).
24. Fu, Y. *et al.* Differential genome-wide profiling of tandem 3′ UTRs among human breast cancer and normal cells by high-throughput sequencing. *Genome Res.* **21,** 741–747 (2011).
25. Ameur, A. *et al.* Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nature Struct. Mol. Biol.* **18,** 1435–1440 (2011).
26. Cole, C. *et al.* Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. *RNA* **15,** 2147–2160 (2009).
27. Kawaji, H. *et al.* Hidden layers of human small RNAs. *BMC Genom.* **9,** 157 (2008).
28. Lee, Y. S., Shibata, Y., Malhotra, A. & Dutta, A. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev.* **23,** 2639–2649 (2009).
29. Park, E., Williams, B., Wold, B. & Mortazavi, A. A Survey of RNA Editing in the human ENCODE RNA-seq data (GRCP043). *Genome Res.* (in the press).
30. Li, M. *et al.* Widespread RNA and DNA sequence differences in the human transcriptome. *Science* **333,** 53–58 (2011).
31. Rozowsky, J. *et al.* AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol. Syst. Biol.* **7,** 522 (2011).
32. Faulkner, G. J. *et al.* The regulated retrotransposon transcriptome of mammalian cells. *Nature Genet.* **41,** 563–571 (2009).
33. Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465,** 182–187 (2010).
34. Ren, B. Transcription: Enhancers make non-coding RNA. *Nature* **465,** 173–174 (2010).
35. Wang, D. *et al.* Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474,** 390–394 (2011).
36. Yip, K. Y. *et al.* Classification of human genomic regions based on experimentally-determined binding sites of more than 100 transcription-related factors. *Genome Biol.* (in the press).
37. Hoffman, M. *et al.* Integrative annotation of chromatin elements from ENCODE data. *Genome Res.* (in the press).
38. Arvey, A., Agius, P., Noble, W. S. & Leslie, C. Sequence and chromatin determinants of cell-type specific transcription factor binding. *Genome Res.* (in the press).
39. Kundaje, A. Ubiquitous heterogeneity and asymmetry of the chromatin landscape at transcription regulatory elements. *Genome Res.* (in the press).
40. Miller, B. Pre-programming of chromatin structure across the cell cycle. *Genome Res.* (in the press).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was supported by the National Human Genome Research Institute (NHGRI) production grants U54HG004557, U54HG004555, U54HG004576 and U54HG004558, and by the NHGRI pilot grant R01HG003700. It was also supported by the NHGRI ARRA stimulus grant 1RC2HG005591, the National Science Foundation (SNF) grant 127375, the European Research Council (ERC) grant 249968, a research grant for the RIKEN Omics Science Center from the Japanese Ministry of Education, Culture, Sports, Science and Technology, and grants BIO2011-26205, CSD2007-00050 and INB GNV-1 from the Spanish Ministry of Science. We would also like to thank C. Gunter and W. Spitzer for editorial assistance with the manuscript.

**Author Contributions** T.R.G., R.G., P.C., B.W., Y.R., M.C.G., G.H., S.E.A., A.R., T.H., M.G. and Y.H. led the project and oversaw the analysis. C.A.D., X.R., B.A.W. and P.C. oversaw or significantly contributed to data generation. S.D., A.Me., A.D., T.L., A.Mo., A.T., J.L., W.L., F.S., C.X., G.K.M., J.K., C.Z., J.R., M.R., F.K. and J.H. made major contributions towards data processing and analysis. R.F.A., T.A., I.A., M.T.B., N.S.B., P.B., K.B., I.B., S.C., X.C., J.Ch., J.Cu., T.D., J.Dr., E.D., J.Du., R.D., E.F., M.F., K.F-T., P.F., S.F., M.J.F., H.Ga., D.G., A.G., H.Gu., C.H., S.J., R.J., P.K., B.K., C.K., O.J.L., E.P., K.P., J.B.P., P.R., B.R., D.R., M.S., L.S., L.-H.S., A.S., J.S., A.M.S., H.Ta., H.Ti., D.T., N.W., H.W., J.W. and Y.Y. were responsible for data production and analysis. T.R.G. and R.G. wrote the manuscript with input from all authors.

**Author Information** A complete set of data files can be downloaded at GEO under accession codes GSE26284 (CSHL, long RNA), GSE33480 (Caltech, A+ RNA-seq), GSE24565 (CSHL, short RNA), GSE33600 (GIS, RNA-PET) and GSE34448 (RIKEN, CAGE) or are viewable at the UCSC Genome Browser (http://genome-preview.ucsc.edu/ENCODE/). Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and the online version of the paper is freely available to all readers. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.R.G. (gingeras@cshl.edu) or R.G. (roderic.guigo@crg.eu).

doi:10.1038/nature11279

# The long-range interaction landscape of gene promoters

Amartya Sanyal[1]*, Bryan R. Lajoie[1]*, Gaurav Jain[1] & Job Dekker[1]

**The vast non-coding portion of the human genome is full of functional elements and disease-causing regulatory variants. The principles defining the relationships between these elements and distal target genes remain unknown. Promoters and distal elements can engage in looping interactions that have been implicated in gene regulation[1]. Here we have applied chromosome conformation capture carbon copy (5C[2]) to interrogate comprehensively interactions between transcription start sites (TSSs) and distal elements in 1% of the human genome representing the ENCODE pilot project regions[3]. 5C maps were generated for GM12878, K562 and HeLa-S3 cells and results were integrated with data from the ENCODE consortium[4]. In each cell line we discovered $>$1,000 long-range interactions between promoters and distal sites that include elements resembling enhancers, promoters and CTCF-bound sites. We observed significant correlations between gene expression, promoter–enhancer interactions and the presence of enhancer RNAs. Long-range interactions show marked asymmetry with a bias for interactions with elements located $\sim$120 kilobases upstream of the TSS. Long-range interactions are often not blocked by sites bound by CTCF and cohesin, indicating that many of these sites do not demarcate physically insulated gene domains. Furthermore, only $\sim$7% of looping interactions are with the nearest gene, indicating that genomic proximity is not a simple predictor for long-range interactions. Finally, promoters and distal elements are engaged in multiple long-range interactions to form complex networks. Our results start to place genes and regulatory elements in three-dimensional context, revealing their functional relationships.**

Spatial proximity and specific long-range interactions between genomic elements can be detected using chromosome conformation capture (3C)-based methods[5]. Previous studies have been limited to analysis of single loci[5–8], interactions that involve a single protein of interest[9], or to analysis of genome-wide folding of chromosomes at a resolution that cannot detect specific looping interactions between genes and functional elements[10]. To overcome these limitations we previously developed 5C (ref. 2). 5C is a high-throughput adaptation of 3C and uses pools of reverse and forward 5C primers to detect long-range interactions between two targeted sets of genomic loci, for example, promoters and distal gene regulatory elements in this study. By targeting a specific part of the genome, 5C facilitates detection of interactions at single restriction fragment resolution.

To begin to define the principles of long-range gene regulation in the human genome we have used 5C to map interactions systematically between promoters and distal elements throughout the 44 ENCODE pilot project regions representing 1% (30 megabases (Mb), Supplementary Table 1) of the genome in three cell lines (Fig. 1a). The ENCODE regions, ranging in size from 500 kilobases (kb) to 1.9 Mb, were selected for comprehensive annotation by the ENCODE pilot project[11]. Here we analysed interactions between 628 TSS-containing restriction fragments and 4,535 'distal' restriction fragments covering the ENCODE regions (Fig. 1a and Supplementary Tables 2 and 3; see also Methods).

**ENCODE**
Encyclopedia of DNA Elements
nature.com/encode

5C libraries were generated for two biological replicates of GM12878, K562 and HeLa-S3 (Supplementary Tables 4–6). These cell lines are extensively annotated by the ENCODE consortium[3,4]. 5C interaction frequencies measured between ENCODE regions located on different chromosomes were used to quantify minor variations in interaction detection efficiencies due to technical biases related to 5C primer efficiency, restriction fragment length, or digestion efficiency. 5C interaction frequencies were then corrected for these biases (Methods and Supplementary Data).

An example of a 5C long-range interaction map representing TSS–distal fragment interactions along and between 14 ENCODE regions (ENm001–ENm014) is shown in Fig. 1b. 5C detects known general features of spatial chromatin organization. First, interactions within the same ENCODE region are more frequent than those between different ENCODE regions. Within one ENCODE region interaction frequencies are generally higher for pairs of loci located closer together in the linear genome. This inverse relationship between genomic distance and interaction frequency is as expected for a flexible chromatin fibre[5,12]. Second, interactions between ENCODE regions that are located on the same chromosome are more frequent than interactions between regions located on different chromosomes (arrow in Fig. 1b). This is consistent with 4C and Hi-C analyses[6,10], and is due to the formation of spatially separated chromosome territories.

5C data sets were analysed to identify TSS–distal fragment pairs that interact more frequently than expected, indicating that they are relatively close in space. For each biological replicate we independently determined the average relationship between interaction frequency and genomic distance (solid red lines in Fig. 1c, d). We defined this as the expected interaction frequency. Next we identified interactions that occur significantly more frequently than expected for loci separated by a corresponding genomic distance by transforming 5C signals into a *z*-score (false discovery rate (FDR) = 1%; Methods). Specific long-range interactions are then defined as pairs of loci that interact significantly more frequently than expected in both replicates. By excluding interactions that are significant in only one replicate, we estimate that only around 10–18% of the significant long-range interactions identified by our approach might be false positives, as estimated from analysis of interactions in gene desert ENCODE regions (ENr112, ENr113 and ENr313) where no significant long-range interactions were expected (Methods). This application of stringent thresholds probably leads to a higher false-negative rate. Consistently, interaction frequencies that are found to be significant in only one replicate are still significantly elevated in the other replicate as

---

[1] Program in Systems Biology, Program in Gene Function and Expression, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 364 Plantation Street, Worcester, Massachusetts 01605-0103, USA.

*These authors contributed equally to this work.

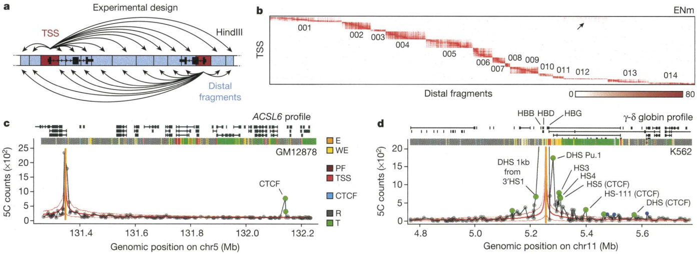


**Figure 1 | 5C approach to identify looping interactions. a,** 5C design[28]. Reverse 5C primers were designed for HindIII fragments that contain a TSS (red; according to the GENCODE v7[20]) and forward 5C primers for all other 'distal' HindIII fragments (blue). **b,** Heat map of all interrogated TSS–distal fragment interactions in 14 ENCODE regions (ENm001–ENm014) in K562 cells. Fragments are displayed in their genomic order. Each dark rectangular area in the heat map denotes interactions within a single ENCODE region whereas remaining areas denote interactions between regions. ENCODE regions that are on the same chromosome show a higher interaction frequency (arrow) than regions that were on different chromosomes. **c, d,** Examples of 5C interaction profiles for two TSSs indicated by vertical orange bars (left, *ACSL6* gene located in ENm002; right, γ-δ-globin located in ENm009). The solid red lines show the expected interaction level (Lowess line, Methods); dashed red lines above and below indicate Lowess ± 1 standard deviation. 5C signals that are significantly higher than expected in both biological replicates (green circles, FDR = 1%) are considered looping interactions. Interactions that are significant in only one replicate (blue circles) are not considered as a high-confidence 5C looping interaction. 5C peak calling detects a long-range interaction between the TSS of *ACSL6* and a distal CTCF-bound element in GM12878 cells. The approach identifies the known long-range interactions of γ-δ-globin to HS3, HS4, HS5 and HS-111 and several additional DHS and CTCF sites in K562 cells[2] (labelled).

compared to interactions that are never significant, but are just below the chosen 1% FDR threshold (Supplementary Fig. 1).

Our analysis correctly identified known interactions between TSSs and their cognate distal regulatory elements, providing validation of the approach (Supplementary Fig. 3). As an example, Fig. 1d shows the 5C interaction profile in K562 cells for a TSS located in the β-globin locus. We previously found that this TSS located just downstream of the γ-globin genes displayed prominent looping interactions with the distal locus control region (LCR) in K562 cells[2]. Our analysis accurately detected these looping interactions (HS3, HS4 and HS5). We identified additional known long-range interactions with DNase I hypersensitive sites (DHSs) near distal CTCF-bound elements (3′HS1 and HS-111)[2,13,14]. In K562 cells we also detected the known interactions between the γ-globin gene (*HBG1*) and the LCR (HS5) and between the α-globin genes and three distal regulatory elements including the α-globin enhancer HS40, and two CTCF-bound elements (HS46 and HS10), located 40, 46 and 10 kb upstream of the genes, respectively (Supplementary Fig. 3 and refs 15, 16). The importance of these distal elements in regulating globin gene expression through looping has been extensively documented[14,16]. As expected, these looping interactions in the globin loci were not detected in GM12878 or HeLa-S3 cells that express little or no globin (Supplementary Fig. 3). Additional examples of cell-type-specific TSS–distal element interactions are shown in Supplementary Fig. 4. Furthermore, 5C interaction frequencies are correlated with TSS–distal DHS pairs predicted to be functionally connected based on their highly correlated activity across a large panel of cell lines ($P < 10^{-13}$, one-sided Mann–Whitney *U*-test[17]), providing independent validation of their biological significance.

In each cell line we identified large numbers of statistically significant TSS–distal fragment interactions, of which ~60% were observed in only one of the three cell lines (Fig. 2a). These data point to intricate cell-type-specific three-dimensional folding of chromatin. 3C-based assays detect specific and functional interactions, for example, TSSs with gene regulatory elements[8]. In addition, the assay will detect 'structural' interactions, for example, close spatial proximity as a result of other nearby specific looping interactions (bystander interactions) or overall higher order folding of the chromatin fibre. To determine which looping interactions involved distal sites that displayed specific chromatin features associated with functional elements, we compared our data with data sets generated by the ENCODE consortium (Fig. 2b and Supplementary Table 7). We found that looping interactions in all cell lines were significantly enriched for distal fragments that are bound by CTCF—a protein known to mediate DNA looping[18]—contain open chromatin (as determined by FAIRE[19] or DHS mapping[17]), and/or contain histones with modifications that are characteristic for active functional elements (H3K4me1, H3K4me2 and H3K4me3). Long-range interactions are also enriched for H3K9ac and H3K27ac, but are not enriched or significantly depleted for H3K27me3, a mark typically found at inactive or closed chromatin.

To gain more insight into the types of element present in the distal looping fragments, we made use of genome-wide and cell-line-specific segmentation analyses that identified seven distinct chromatin states based on histone modifications, the presence of DHSs and the localization of proteins such as RNA polymerase II and CTCF (ref. 4 and Fig. 2b). These states are: (1) enhancer (E); (2) weak enhancer (WE); (3) TSS; (4) predicted promoter flanking regions (PF); (5) insulator element (CTCF); (6) predicted repressed region (R); and (7) predicted transcribed region (T). The ENCODE consortium tested sets of the E elements in enhancer assays and confirmed that >50% display enhancer activity[4]. We found that looping interactions were significantly enriched for distal fragments that contained E, WE and CTCF elements, and the actively transcribed chromatin state (T), but were depleted for the repressed chromatin state (R). We note that some distal looping fragments contained elements classified as TSS or PF, even though they did not contain TSSs as defined by the GENCODE v7 annotation[20]. Possibly, these are yet-to-be-annotated TSSs.

Next, we used the seven-way segmentation data to categorize looping interactions into four broader functional groups (Fig. 2c, Supplementary Fig. 5 and Supplementary Data): those that involve a distal fragment that contains a putative enhancer ('E' (E or WE)), a putative promoter ('P' (TSS or PF)), or a CTCF-bound element

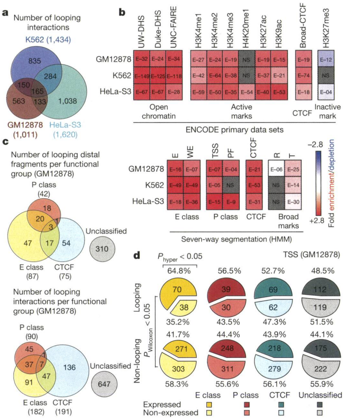


**Figure 2 | Distribution of looping interactions across cell types and their relationship with chromatin features and gene expression.** **a**, Venn diagram showing the number of unique and overlapping looping interactions across three cell types. **b**, Heat map showing the enrichment/depletion of chromatin features in looping fragments compared to all interrogated fragments based on genome-wide data sets from the ENCODE consortium (Supplementary Table 7). Features include open chromatin (UW-DHS (UW, University of Washington), Duke-DHS and UNC-FAIRE (UNC, University of North Carolina; FAIRE, formaldehyde-assisted isolation of regulatory elements)); active marks (Broad Institute histone H3K4me1/2/3, H4K20me1, H3K27ac, H3K9ac); CTCF (Broad Institute CTCF ChIP peaks); inactive marks (Broad Institute histone H3K27me3); and seven-way segmentation[4] (based on HMM prediction for indicated cells). We further grouped segmentation categories E and WE into 'E class', TSS and PF into 'P class', and R and T into 'broad marks'. The colour scale represents the fold enrichment (red) or depletion (blue). The numbers listed inside each box represent $P$ values of the significant ($P < 0.05$) enrichment/depletion for that mark, where (for example) E−32 indicates $\times 10^{-32}$ (NS, not significant, grey; two-tailed hypergeometric test and corrected for multiple testing using Bonferroni). **c**, Venn diagram showing the number of unique and overlapping looping distal fragments (top) and looping interactions (bottom) among four functional groups in GM12878 cells. Distal fragments are classified into four non-exclusive groups based on the seven-way segmentation. Similarly, TSS–distal fragment interactions are classified based on the functional grouping of the distal fragments. The four functional groups are E class (yellow), P class (magenta), CTCF (cyan) and unclassified (grey). **d**, Pie charts showing percentages and numbers of expressed/non-expressed TSSs looping or not looping to a particular group (E, P, CTCF or unclassified; coloured as in **c**) of distal fragments in GM12878 cells. TSSs with a CAGE value >0 are deemed expressed. Significant enrichment for expressed TSSs in the looping or non-looping categories is indicated on top (hypergeometric test; $P_{\mathrm{hyper}} < 0.05$). Significant differences in expression levels between TSS in the looping versus the non-looping category is indicated on the left (Wilcoxon signed-rank test; $P_{\mathrm{Wilcoxon}} < 0.05$).

(CTCF). The final class contains interactions with fragments that do not contain any of these three types of element, although they do contain T and R states ('U', unclassified). The last class is relatively large but is still significantly enriched in features that are characteristic for active functional elements such as H3K4me1, and over 60% of the unclassified fragments contain chromatin features found at active chromatin elements (Supplementary Fig. 7). Thus, these are not simply noise or false positives, but are probably the result of the conservative segmentation approach.

We found that TSS–E and TSS–P interactions are more cell-type specific than TSS–CTCF interactions: for the TSS–E and TSS–P categories, the ratio of interactions that is seen in only one cell line versus more than one cell line is ~4:1, whereas it is close to ~1:1 for the TSS–CTCF category (Supplementary Fig. 5). The cell-type-specific activity of some of these E elements was confirmed using transient reporter assays (Supplementary Fig. 10). Next, we determined whether looping of a TSS to any of the four categories of chromatin states is correlated with transcription. We used CAGE expression data[21] to assign an expression level to each TSS. We found that looping interactions with fragments containing enhancer-like E elements were significantly enriched for those that involved expressed TSSs (Fig. 2d and Supplementary Fig. 6). In addition, the subset of TSSs that interact with fragments containing E elements was significantly more highly expressed compared to TSSs that do not interact with E elements. Interactions with other classes of element (CTCF, P and U) are significantly enriched for actively expressed genes in some, but not all, cell lines (Supplementary Fig. 6).

Active enhancers often express enhancer RNAs[22]. We used a comprehensive enhancer RNA data set generated by the ENCODE consortium to determine whether TSSs preferentially interact with active enhancer-like elements[23]. We found that E elements that are looping to TSSs are significantly more likely to express enhancer RNAs than E elements that are not looping ($P < 5 \times 10^{-5}$, hypergeometric test, Supplementary Fig. 10). We conclude that looping interactions preferentially involve active enhancer-like elements.

Next we analysed the distribution of long-range interactions upstream and downstream of TSSs. To generate this landscape of looping interactions we aligned all TSSs and calculated the average number of interactions that a TSS has with each class of distal element at increasing genomic distances upstream and downstream of the TSS. Figure 3a shows the resulting average long-range interaction profile across all three cell lines (similar results were obtained when each of the cell lines was analysed separately; Supplementary Fig. 8). Notably, we found that the long-range interaction landscape is asymmetric, with interactions of E, P and CTCF classes peaking around 120 kb upstream of the TSS. This asymmetry of interactions reveals an unanticipated directionality in long-range interactions with TSSs. This may indicate the presence of topological constraints imposed by the mechanism by which such interactions regulate target promoters. No such bias was observed for the set of unclassified elements, or for the complete set of interrogated interactions (Fig. 3a). Interestingly, previous analyses showed that conserved non-coding elements are also often found within similar distances of target genes[24]. Third, when we analysed expressed TSSs and non-expressed TSSs separately, we found that both have a similar interaction landscape but that expressed TSSs tend to have more interactions, especially with the E, P and CTCF classes. We cannot rule out the possibility that some TSSs classified as non-expressed based on the absence of CAGE tags are actually expressed at low levels.

Next we explored whether the relative order of elements in the genome affects which long-range interactions occur. It is often assumed that distal elements such as enhancers target the nearest TSS. Only ~7% of the looping interactions are between an element and the nearest TSS (Fig. 3b). This number goes up to 22% when only active TSSs are included. Similarly, 27% of the distal elements have an interaction with the nearest TSS, and 47% of elements have interactions with the nearest expressed TSS. Thus, when predicting TSS–distal element interactions, choosing the nearest (active) gene is often not correct.

It has been suggested that CTCF sites located between an enhancer and a TSS may prevent enhancer–promoter interactions[18,25], although in individual cases interactions over such sites have been observed[14,26].

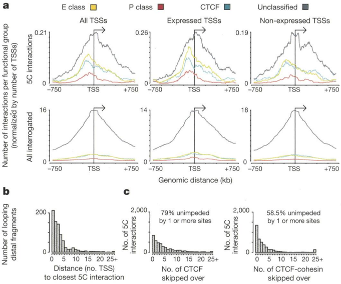


**Figure 3 | Looping landscape of TSSs to distal fragments.** **a**, Composite profile of average number of group-specific looping interactions upstream and downstream of TSSs on the basis of combined 5C interaction data from the three cell lines. The top panel shows the average looping profiles of all TSSs (left), of expressed TSSs (middle) and of non-expressed TSSs (right). The bottom set of plots shows the corresponding profiles of all interrogated TSS–distal element interactions (left), of expressed TSSs (middle) and of non-expressed TSSs (right). All the interaction data for a particular group for all three cell lines are binned with a sliding window of 150 kb (step size of 5 kb) and normalized for the number of TSSs. **b**, Histogram showing the number of distal fragments that are involved in looping with their target promoters skipping 0,1,2,…,25 (and above) TSSs. **c**, Histogram showing the number of looping interactions that skip over 0, 1, 2,…, 25 (and above) restriction fragments bound by either CTCF (left) or by both CTCF and RAD21 (cohesin; right). In **b** and **c** combined results for all three cell lines are plotted and values above 24 on the *x* axis are added and grouped as 25+. Percentage of looping interactions that skip $\geq 1$ CTCF (left) or CTCF plus cohesin (right) are indicated on top.

To address this question we determined the frequency of identified long-range interactions between a TSS and a distal element that skip over one or more sites bound by CTCF. We found that 79% of long-range interactions are unimpeded by the presence of one or more CTCF-bound sites (Fig. 3c). Thus, the presence of a CTCF-bound site does not block physical long-range interactions. It has been reported that CTCF acts in conjunction with the cohesin complex to block promoter–enhancer interactions[27]. We found that 58% of looping interactions skip sites co-bound by CTCF and cohesin (Fig. 3c). We obtained similar results when the different categories of long-range interaction (TSS–E, TSS–P, TSS–CTCF and TSS–U) were analysed separately. Possibly, additional factors need to be recruited to CTCF-bound sites to acquire interaction-blocking activity.

The large number of long-range interactions that we discovered indicate that distal elements and TSSs are each engaged in multiple long-range interactions. To characterize this phenomenon in more detail we determined the interaction degree of TSSs and distal fragments. We found that ~50% of TSSs display one or more long-range interaction, with some interacting with as many as 20 distal fragments (Fig. 4a). Expressed TSSs interact with slightly more fragments as compared to non-expressed TSSs (the mean for GM12878 is 1.88 versus 1.37, or 3.88 versus 3.25 when including only those TSSs with at least one interaction). Out of all distal fragments interrogated, ~10% interacted with one or more TSS, with some interacting with more than 10 (mean of 2.15 (for GM12878) when including only those distal fragments with at least one interaction). The degree distribution of the four categories of distal elements was very similar (Supplementary Fig. 9).

Figure 4b shows an example of the complex long-range interaction networks formed by TSSs and distal fragments in the ENr132 region in K562 cells. It is unlikely that these interactions can all occur at the same time in the same cell, which is indicative of significant cell-to-cell variation. The data indicate that gene–element interactions are not exclusively one-to-one, and suggest that multiple genes and distal elements can assemble in larger clusters, as proposed for the β-globin locus[14].

Our data provide new insights into the landscape of chromatin looping that bring genes and distant elements in close spatial proximity. In addition to generating a rich data set reflecting specific gene–element interactions, the average interaction profile of TSSs with surrounding chromatin reveals several general principles regarding the asymmetric relationships between genomic distance, the order of elements, and the formation of looping interactions. The bias for upstream interactions may indicate that the protein complexes on many TSSs may be asymmetric and may preferentially interact on one side with enhancer–protein complexes. It is also possible that the asymmetry of the long-range interaction landscape reflects a potential preference of looping to elements that are located in intergenic non-transcribed regions. Furthermore, although these average long-range interaction landscapes may facilitate computational prediction of long-range interactions throughout the genome, the fact that interactions skip genes and CTCF/cohesin sites indicates that additional mechanisms for target selection and gene insulation exist.

Although conventional 3C may still be the method of choice to study the folding of individual loci, the 5C design strategy and data analysis methods applied here may provide a general approach for systematically

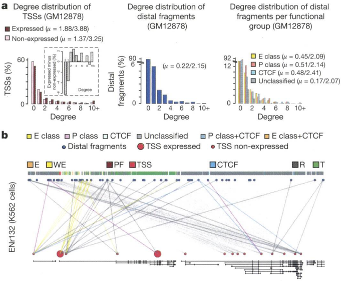


**Figure 4 | Networks of looping interactions.** **a**, Histogram showing the number of TSSs (left, red) or distal fragments (middle, blue) in percentages that are involved in 0, 1, 2,….10 (and above) looping interactions (degree, *x* axis) in GM12878 cells. All of the values for degrees that are $>9$ are grouped under degree 10+. The dark red bars represent the percentages of looping TSSs that are expressed whereas light red bars represent the percentages of looping TSSs that are not expressed. Inset: the difference in percentage between looping TSSs that are expressed and not expressed for each degree is shown. The right panel shows the degree distribution for each functional group of distal fragments. The average degrees (mean, $\mu$) for TSSs and distal fragments are indicated. The first value is the mean degree considering all the TSS/distal fragments (looping plus non-looping), whereas the second value is the mean degree of looping TSS/distal fragments (excluding degree = 0). **b**, Web plot showing the long-range looping interactions in the ENr132 region in K562 cells. The interrogated distal fragments (blue circles) and the TSSs (red circles) are positioned according to genomic coordinates and the GENCODE v7 gene annotation is indicated. The size of the red circles indicates whether that TSS is expressed (large circles) or not expressed (small circles). The thin grey lines show all the interactions that were interrogated. The coloured lines show significant looping interactions between TSSs and distal fragments of a particular group.

mapping gene–element interactions for large gene sets. With further development of 3C technology and increases in sequencing capacity, similar high-resolution studies should become feasible to map specific long-range interactions throughout the genome, which may uncover additional principles that guide chromatin looping. Such insights will also be critical for interpreting genome-wide association studies that often identify regions with regulatory elements but not their distally located target genes. Co-published ENCODE-related papers can be explored online via the *Nature* ENCODE explorer (http://www.nature.com/ENCODE), a specially designed visualization tool that allows users to access the linked papers and investigate topics that are discussed in multiple papers via thematically organized threads.

## METHODS SUMMARY

5C was performed using two pools of 5C primers: one for ENm001–ENm014 and ENr313, and one pool for all 30 randomly picked ENCODE regions (ENr111–ENr334)[11] (Supplementary Tables 2 and 3). 5C libraries (two biological replicates per cell line) were sequenced on an Illumina GAIIx platform and sequence reads were mapped using Novoalign (http://www.novocraft.com), as described[15]. Interaction data for each experiment are available in GEO (accession number GSE39510). Statistically significant pair-wise interactions were identified (Methods) by converting each 5C signal into a $z$-score using the average 5C signal distribution versus genomic distance as a background estimate. Significant interactions (1% FDR) observed in both biological replicates were considered looping interactions. 5C looping interactions were compared to a variety of genome-wide data sets generated by the ENCODE consortium[4] (Supplementary Table 7).

**Full Methods** and any associated references are available in the online version of the paper.

**Received 9 December 2011; accepted 1 June 2012.**

1. Dekker, J. Gene regulation in the third dimension. *Science* **319,** 1793–1794 (2008).
2. Dostie, J. *et al.* Chromosome conformation capture carbon copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Res.* **16,** 1299–1309 (2006).
3. ENCODE Project Consortium. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* **9,** e1001046 (2011).
4. ENCODE Project Consorium. An integrated encyclopedia of DNA elements in the human genome. *Nature* http://dx.doi.org/10.1038/nature11247 (this issue).
5. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295,** 1306–1311 (2002).
6. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature Genet.* **38,** 1348–1354 (2006).
7. Zhao, Z. *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature Genet.* **38,** 1341–1347 (2006).
8. Miele, A. & Dekker, J. Long-range chromosomal interactions and gene regulation. *Mol. Biosyst.* **4,** 1046–1057 (2008).
9. Fullwood, M. J. *et al.* An oestrogen-receptor-α-bound human chromatin interactome. *Nature* **462,** 58–64 (2009).
10. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326,** 289–293 (2009).
11. ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447,** 799–816 (2007).
12. Gheldof, N., Tabuchi, T. M. & Dekker, J. The active FMR1 promoter is associated with a large domain of altered chromatin conformation with embedded local histone modifications. *Proc. Natl Acad. Sci. USA* **103,** 12463–12468 (2006).
13. Palstra, R. J. *et al.* The β-globin nuclear compartment in development and erythroid differentiation. *Nature Genet.* **35,** 190–194 (2003).
14. Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. & de Laat, W. Looping and interaction between hypersensitive sites in the active β-globin locus. *Mol. Cell* **10,** 1453–1465 (2002).
15. Baù, D. *et al.* The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules. *Nature Struct. Mol. Biol.* **18,** 107–114 (2011).
16. Vernimmen, D., De Gobbi, M., Sloane-Stanley, J. A., Wood, W. G. & Higgs, D. R. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J.* **26,** 2041–2051 (2007).
17. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* http://dx.doi.org/10.1038/nature11232 (this issue).
18. Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137,** 1194–1211 (2009).
19. Song, L. *et al.* Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res.* **21,** 1757–1767 (2011).
20. Harrow, J. *et al.* GENCODE: The reference human genome annotation for the ENCODE project. *Genome Res.* http://dx.doi.org/10.1101/gr.135350.111 (2012).
21. Dong, X. *et al.* Correlating histone modifications and gene expression. *Genome Biol.*. (in the press).
22. Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465,** 182–187 (2010).
23. Djebali, S. *et al.* Landscape of transcription in human cell lines. *Nature* http://dx.doi.org/10.1038/nature11233 (this issue).
24. Vavouri, T., McEwen, G. K., Woolfe, A., Gilks, W. R. & Elgar, G. Defining a genomic radius for long-range enhancer action: duplicated conserved non-coding elements hold the key. *Trends Genet.* **22,** 5–10 (2006).
25. Wallace, J. A. & Felsenfeld, G. We gather together: insulators and genome organization. *Curr. Opin. Genet. Dev.* **17,** 400–407 (2007).
26. Kurukuti, S. *et al.* CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc. Natl Acad. Sci. USA* **103,** 10684–10689 (2006).
27. Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451,** 796–801 (2008).
28. Lajoie, B. R., van Berkum, N. L., Sanyal, A. & Dekker, J. My5C: web tools for chromosome conformation capture studies. *Nature Methods* **6,** 690–691 (2009).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank the University of Massachusetts Medical School Deep Sequencing core for sequencing 5C libraries, and R. Thurman and J. Stamatoyannopoulos for discussion and help with peak calling analysis. We thank M. Walhout and members of the Dekker laboratory for discussions. This work was supported by grants from the National Institutes of Health, National Human Genome Research Institute (HG003143 and HG003143-06S1) and a W.M Keck Foundation Distinguished Young scholar in Medical Research award to J.D.

**Author Contributions** J.D. conceived the project. A.S. performed all experiments. B.R.L. designed 5C experiments, and built the data analysis and visualization pipelines. B.R.L., A.S., G.J. and J.D. analysed the data and wrote the paper.

**Author Information** All data are publicly available at GEO (accession number GSE39510). 5C data has also been deposited in the public UCSC ENCODE database (http://encodeproject.org/ENCODE/). 5C data can be found at http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeUmassDekker5C/. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and the online version of the paper is freely available to all readers. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.D. (job.dekker@umassmed.edu).

nature
nature

A key thread in your scientific career

Since 1869 we've made it our mission to share the greatest developments in science and advance the discussion among the global scientific community. Whether in print, online or mobile, *Nature* is always at the forefront of science and is your forum to read, watch, listen and engage with key research, news and opinion.

**How will you access the next landmark discovery?**

**NATURE.COM/NATURE**

nature publishing group npg

Seek to break new ground.
HiSeq 2500
HiScan
SQ
MiSeq
illumina

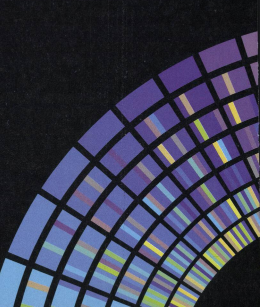
HiSeq® 2000 series HiScan®SQ MiSeq™

Whether you are looking to push limits, move mountains, or inspire hope, we are there to help you move forward faster. For all you seek—Illumina's sequencing solutions will empower you to find it.

For all you seek.

www.illumina.com/CompareSequencers

illumina®



nature

www.nature.com/encode

doi:10.1038/nature11376

# Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis

Anna Plechanovová[1], Ellis G. Jaffray[1], Michael H. Tatham[1], James H. Naismith[2] & Ronald T. Hay[1]

**Ubiquitin modification is mediated by a large family of specificity determining ubiquitin E3 ligases. To facilitate ubiquitin transfer, RING E3 ligases bind both substrate and a ubiquitin E2 conjugating enzyme linked to ubiquitin via a thioester bond, but the mechanism of transfer has remained elusive. Here we report the crystal structure of the dimeric RING domain of rat RNF4 in complex with E2 (UbcH5A) linked by an isopeptide bond to ubiquitin. While the E2 contacts a single protomer of the RING, ubiquitin is folded back onto the E2 by contacts from both RING protomers. The carboxy-terminal tail of ubiquitin is locked into an active site groove on the E2 by an intricate network of interactions, resulting in changes at the E2 active site. This arrangement is primed for catalysis as it can deprotonate the incoming substrate lysine residue and stabilize the consequent tetrahedral transition-state intermediate.**

By altering the fate of modified proteins, conjugation with ubiquitin and its homologues has a central role in eukaryotic biology underpinning cell signalling, protein degradation and stress responses. In most cases ubiquitin is transferred to its target proteins from a thioester complex with a ubiquitin conjugating enzyme (E2) by a large family of ubiquitin E3 ligases (E3)[1]. The RING family of E3s, of which over 600 are encoded in the human genome, possess a conserved arrangement of cysteine and histidine residues that coordinate two zinc atoms[2]. RING E3 ligases bind both substrate and E2–ubiquitin (E2–Ub) thioester, but the molecular basis by which the RING activates the E2–Ub bond for transfer of ubiquitin to substrate has remained elusive.

RNF4 is a SUMO-targeted ubiquitin ligase[3] that has a key role in the DNA damage response[4–6] and in arsenic therapy for acute promyelocytic leukaemia[7,8]. RNF4 contains multiple SUMO interaction motifs, allowing it to engage polySUMO-modified substrates, and a RING domain[2] that is responsible for dimerization and catalysis of ubiquitin transfer[3,9]. Our understanding of RING-catalysed ubiquitination has been hindered by the lack of structures of the key intermediate: a RING bound to E2–Ub. Obtaining this key complex is difficult, as the thioester (or engineered oxyester) bond linking E2 and ubiquitin is highly activated and unstable in the presence of an E3.

## Structure of the RING–UbcH5A–Ub complex

We have engineered a mimic of the E2–Ub thioester bond by replacing the active site cysteine of the E2 UbcH5A (also called UBE2D1) with a lysine to generate an isopeptide (amide) bond between the C terminus of ubiquitin and the ε-amino group of the introduced lysine (Supplementary Figs 1 and 2). Isopeptide-linked UbcH5A–Ub bound selectively to the RNF4 RING and acted as a potent inhibitor of RNF4-mediated substrate ubiquitination, confirming that it is an excellent mimic, but crucially, that it is stable in the presence of RNF4 (Supplementary Fig. 3). The E2–Ub mimic was mixed in a 2:1 ratio with a fused RNF4 RING dimer[3] and crystallized. A 2.2 Å structure of the resulting complex was determined (Supplementary Table 1). The asymmetric unit contains the central RNF4 RING dimer, two UbcH5A molecules and two ubiquitin molecules related by a two-fold axis (Fig. 1). Each UbcH5A molecule contacts a single RING domain and is linked by an isopeptide bond to ubiquitin (Supplementary Fig. 4) that sits at the RING dimer interface. The complex can be envisaged as a dimer of heterotrimers (RING monomer, UbcH5A and ubiquitin).

Strikingly, ubiquitin is folded back onto the E2, creating an interface that buries approximately 1,800 Å$^2$, has 15 hydrogen bonds and 4 salt bridges. L8 of ubiquitin interacts with L97 and K101 of UbcH5A, whereas I44, H68 and V70 in ubiquitin are close to L104, S105 and S108 on the α2 helix of the E2 (Fig. 2a). Extensive contacts are evident between the C-terminal 6 residues of ubiquitin and loops surrounding the active site of UbcH5A, particularly residues L86, D87, Q92 and N114. The side chain of N77 in UbcH5A forms a hydrogen bond to the isopeptide carbonyl (Fig. 2b). Mapping conserved E2 residues (Supplementary Fig. 5) shows that highly conserved residues surround the active site and the shallow groove that accommodates the C-terminal region of the linked ubiquitin (Supplementary Fig. 6). The other conserved cluster of E2 residues constitutes the binding site for the E3 ligase.

UbcH5A contacts a single protomer of the RING (Supplementary Fig. 7) and the interface is very similar to that previously described for RING–E2 complexes[10,11]. At the junction of the three molecules in the heterotrimer is a hydrophobic cluster formed by L8, T9 and L71 of ubiquitin, A96 and L97 of UbcH5A, and P137, P178 and R181 of the RING (Fig. 2c and Supplementary Fig. 7). Ubiquitin contacts both protomers of the RING dimer and the interface buries 940 Å$^2$ (Fig. 2c). Residues L8 to K11 and L71 with R72 of ubiquitin contact RING residues T179 to R181 within the same heterotrimer, whereas the Q31 to Q40 region of ubiquitin contacts both protomers of the RING dimer. The backbone carbonyl of ubiquitin E34 makes a hydrogen bond with RING residue H160 (zinc ligand) and the main-chain E34 to G35 of ubiquitin stacks with the side chain of Y193 of the RING domain from the other heterotrimer (Fig. 2d). These interfaces explain why dimerization of the RNF4 RING is required for activity[3,9]. Phylogenetic analysis of RNF4 from a wide range of species and sequence comparison of RNF4 with other dimeric RING and U-box E3 ligases indicate that the bound ubiquitin interacts with conserved features of the RING (Supplementary Fig. 8).

The RING domain does not undergo any major structural change as a result of complex formation (Supplementary Fig. 9a). Ubiquitin

---

[1] Wellcome Trust Centre for Gene Regulation and Expression, College of Life Sciences, University of Dundee, Dundee DD1 5EH, UK. [2] Biomedical Sciences Research Complex, University of St Andrews, St Andrews KY16 9ST, UK.

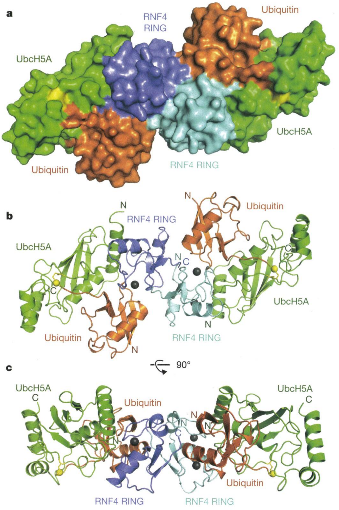


**Figure 1 | Structure of the RNF4 RING bound to ubiquitin-loaded UbcH5A.** **a**, Surface representation of the complex. Individual RING protomers are coloured cyan and blue, UbcH5A is green, ubiquitin is orange and the isopeptide linkage between the C terminus of ubiquitin and K85 of UbcH5A is shown in yellow. **b**, Ribbon diagram of the complex with the same orientation and colour scheme as in **a**. Zinc atoms are indicated as grey spheres. **c**, As in **b**, but the complex is rotated by 90° as indicated.

shows little change in overall structure up to R72; the remaining five residues are, however, positioned differently as a consequence of being held in the active site groove of UbcH5A. The loop at L8 in ubiquitin has moved over 4 Å to form the hydrophobic cluster with UbcH5A and the RING (Supplementary Fig. 9b). Superposition of the coordinates of unconjugated UbcH5A either free (Protein Data Bank accession 2C4P), or in a variety of non-covalent complexes[12–14], and UbcH5A in the present structure reveals a clear re-arrangement centred on D117. In the unconjugated structures, the side chain of D117 points towards C85, in a position that would clash with the isopeptide (thioester) bond observed in our complex (Supplementary Fig. 9c, d).

## E2 and ubiquitin residues required for activity

Previous mutational analysis revealed the importance of the RING residue R181—which contacts both E2 and ubiquitin in the present structure (Fig. 2)—in the ubiquitination activity of RNF4 (ref. 3). Moreover, Y193 in the RING plus L8 and I44 in ubiquitin were shown to be required for activity[3]. Although it was thought that these residues might interact directly, the present structure emphasizes their importance but shows that they are not in direct contact. To validate our structure further, we introduced mutations into ubiquitin and UbcH5A (Fig. 3a, b) and tested these in a single-turnover substrate ubiquitination assay. Mutations of hydrophobic residues I44 (ubiquitin) and L104 (UbcH5A) at the interface between ubiquitin and UbcH5A abolished ubiquitination activity, whereas mutations K101A, S108A and D112A in UbcH5A and R42A in ubiquitin reduced activity modestly (Fig. 3c, d and Supplementary Figs 10–13). Ubiquitin mutations G35A and I36A (both at the RING interface) substantially ($>10\times$) reduced activity. Significant reductions in ubiquitination were also observed for mutations of L8 and L71 in ubiquitin and L97 in UbcH5A that form a hydrophobic core at the junction of all the three molecules in the heterotrimer. In the E2 active site groove, mutations N77A and D87A in UbcH5A abolished activity, whereas D117A severely compromised activity. N114A in UbcH5A and R72A, L73A and R74A in ubiquitin displayed modestly reduced activity (Fig. 3c, d and Supplementary Figs 10–13).

To discriminate between residues in ubiquitin and E2 that influence the ability of the substrate lysine to carry out nucleophilic attack on the E2–Ub thioester and those residues involved in activating the E2–Ub bond, we carried out substrate-independent assays that measure the ability of the RNF4 RING to catalyse hydrolysis of an E2–Ub oxyester bond[3] (Fig. 3e, f and Supplementary Figs 14 and 15). Mutations in ubiquitin and UbcH5A that reduced substrate-dependent ubiquitination also reduced oxyester hydrolysis, with the important exception of D117A, which was defective in substrate ubiquitination but retained wild-type levels of oxyester hydrolysis (Fig. 3d, f).

We investigated whether residues in ubiquitin and UbcH5A that are important for RNF4-mediated ubiquitination have a more general role in E3-catalysed transfer. The ubiquitin and UbcH5A mutants were tested in combination with the unrelated U-box E3 ligase CHIP (C terminus of Hsp70-interacting protein) using an autoubiquitination assay. Although there are relatively modest quantitative differences in ubiquitination, the effect of the mutations on CHIP and RNF4 activity is very similar (Fig. 4). Thus, it is likely that a conserved mechanism is used by CHIP and RNF4 to catalyse ubiquitin transfer.

## Mechanism of RING-mediated ubiquitination

Using the isopeptide-linked E2–Ub in our crystal structure, we constructed a model of the E2–Ub thioester by replacing K85 in UbcH5A with a cysteine and minimizing the geometry. The resulting model shows very minor changes: the $S\gamma$ and $C\alpha$ atoms in C85 are shifted 1.0 Å and 0.2 Å from $C\gamma$ and $C\alpha$ atoms of K85, with smaller changes in I84 and L86. In ubiquitin the $C\alpha$ atoms of G76 and G75 have moved 0.5 Å and 0.2 Å, respectively. The carbonyl group of the thioester at G76 has moved 0.6 Å and rotated around 45°, resulting in the hydrogen bond with N77 being extended to 3.6 Å (Fig. 5a, b). Coupled with the mutational analysis and evidence that the isopeptide-linked E2–Ub is a competitive inhibitor of ubiquitination, we conclude that the crystal structure is a relevant model for the key E2–Ub–RING heterotrimeric intermediate.

In the absence of an E3 ligase, the ubiquitin thioester linked to the E2 can adopt a wide range of different conformations that also include a 'folded-back' conformation[15–17]. As free ubiquitin has no detectable affinity for the RNF4 RING we suggest that the initial interaction will be between E2 and the RING. In this encounter, with the E2 bound to one RING protomer, the thioester-linked ubiquitin would be engaged by Y193 of the other RING protomer and folded back to contact the $\alpha2$ helix of UbcH5A, while its C terminus is extended and locked in the active site groove of the E2. This orientates the planar thioester bond such that the ubiquitin G76 thioester carbonyl is in the optimal arrangement for nucleophilic attack by the incoming substrate lysine. This arrangement of the E2 active site was not observed in a UbcH5B–Ub oxyester alone[18] or when a UbcH5B–Ub oxyester is bound to a HECT E3 ligase[19] (Fig. 5c, d). The nucleophilic attack by the substrate lysine would result in formation of a tetrahedral intermediate on the G76 carbonyl carbon. The G76 carbonyl oxygen, with its developing

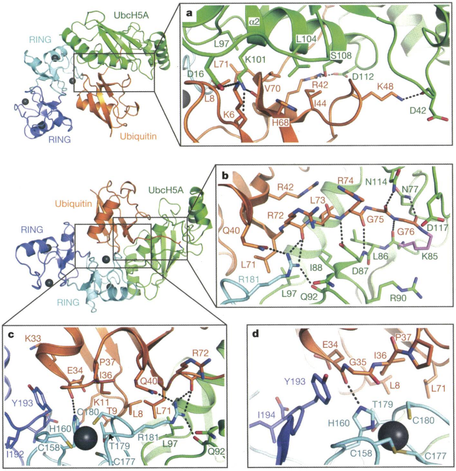


**Figure 2 | Molecular interfaces in the RNF4 RING–UbcH5A–Ub complex.** **a**, Detail of the interaction between ubiquitin (orange) and the α2 helix of UbcH5A (green). **b**, Detail of the interaction interface between ubiquitin (orange) and UbcH5A (green) in the E2 active site groove. The side chain of K85 in UbcH5A that forms the isopeptide bond with ubiquitin is coloured violet. **c**, The hydrophobic cluster at the centre of the ubiquitin (orange), UbcH5A (green) and RING (cyan) heterotrimer. **d**, Stacking interaction between the main chain of ubiquitin (orange) in one heterotrimer and Y193 of the RING (blue) from the other heterotrimer.

negative charge, would move down below the plane of the original thioester bond and form a hydrogen bond to N77, stabilizing the tetrahedral intermediate. In fact the atoms would move towards the experimental orientation of the carbonyl in the isopeptide bond that makes a 2.8 Å hydrogen bond with N77. The role of UbcH5A D117, which sits above the thioester and is re-positioned by ubiquitin binding, has been clarified by analysis of the D117A mutant. Of the mutants which are defective in the ubiquitination assay, only D117A retains wild-type levels of oxyester hydrolysis (Fig. 3f). Because the E2–Ub oxyester bond is hydrolysed in the presence of E3 (no transfer to substrate)[3], only a residue with the sole function to position and/or activate the incoming lysine nucleophile should possess activity in oxyester hydrolysis assays but be inactive in ubiquitination.

## Implications for transfer of ubiquitin and related modifiers

This is the first structure of a RING E3 ligase bound to a ubiquitin-loaded E2, but the mechanism proposed here for ubiquitin transfer to substrate is consistent with previous work. Key roles for residues N77 (ref. 20) and D117 (ref. 21) in E2 catalytic activity have been suggested previously. Evidence that activation of the thioester bond requires both ubiquitin/ubiquitin-like modifier (Ubl) and E2 to be bound by the E3 comes from previous work on RNF4 (ref. 3), the SIZ1 (ref. 22) and RanBP2 (ref. 23) SUMO E3 ligases, and the NEDD4L HECT E3 ligase[19]. The folded-back conformation where the I44 hydrophobic patch of ubiquitin (or equivalent region of SUMO) engages the α2 helix of the E2 has been suggested as an intermediate in ubiquitin/Ubl transfer based on NMR models[15,17], mutagenesis coupled with modelling[22,24], and from the structure of a SUMO substrate–E2–E3 product complex[23,25]. Comparing the NMR model of UBC1 (also called UBE2K)–Ub thioester[15] with the present structure shows that although ubiquitin in the UBC1–Ub thioester is in the folded-back conformation, it is different from the present structure where interactions between ubiquitin and the RING extend and exert tension on the ubiquitin C terminus, locking it down into the E2 active site groove. In the absence of its cognate E3 the ubiquitin C-terminal tail in the UBC1–Ub complex is not locked down in the UBC1 active site groove and the thioester is thus not activated (Supplementary Fig. 16).

The folded-back conformation was also observed in the structure of SUMO-modified RanGAP1 in complex with UBC9 (also called UBE2I) and the SUMO E3 ligase RanBP2 (ref. 25) (trapped product complex). The position of the SUMO C-terminal tail and hydrogen bonding interactions within the active site groove of UBC9 are remarkably similar to those seen for UbcH5A–Ub bound to the RNF4 RING (Supplementary Fig. 17). Although both RNF4 and RanBP2 interact with ubiquitin/SUMO to lock it into this conformation, molecular details of these contacts are rather different. Whereas the RING domain interacts with a hydrophobic patch in ubiquitin containing L8, I36 and L71, RanBP2 holds SUMO using a SUMO interaction motif. Superimposing UBC9 from the RanGAP1–SUMO–UBC9–RanBP2 complex with UbcH5A from the RNF4

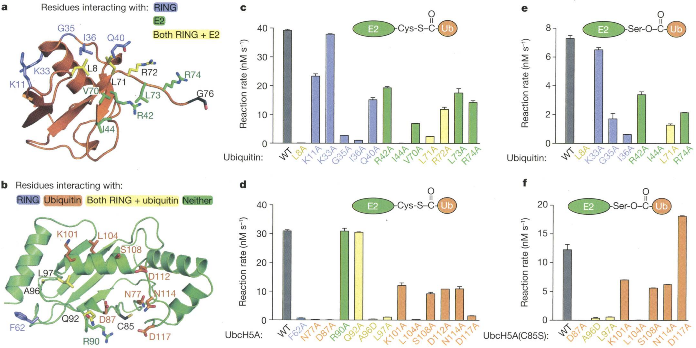


**Figure 3 | Mutational analysis of the RNF4 RING–UbcH5A–Ub complex.** **a**, Side chains of altered residues in ubiquitin contacting RNF4 (blue), UbcH5A (green), or both RNF4 and UbcH5A (yellow). **b**, Side chains of altered residues in UbcH5A contacting RNF4 (blue), ubiquitin (orange), both RNF4 and ubiquitin (yellow), or neither (green). **c**, Reaction rates were determined (mean ± s.d. of duplicates) for single-turnover, RNF4-dependent substrate ubiquitination assays with mutant forms of ubiquitin. Wild-type ubiquitin is in grey and mutants are colour coded as in **a**. **d**, Assays with UbcH5A mutants quantified as in **c** and colour coded as in **b**. **e**, RNF4-mediated hydrolysis of UbcH5A(C85S)–Ub oxyesters with mutations in ubiquitin. Rates are mean ± s.d. of duplicates. **f**, As in **e**, with mutations in UbcH5A.

RING–UbcH5A–Ub structure allows a model of the catalytic transfer complex to be constructed (E2–Ub thioester, E3 and substrate) (Fig. 5e and Supplementary Fig. 17d, e). This model both unifies and provides clear molecular rationale for a body of existing data on ubiquitination.

Although RNF4 is a structurally simple E3 ligase it seems likely that similar principles of E2–Ub activation will be used by structurally more complex ubiquitin ligases such as the cullin-based ligases[26] and the anaphase promoting complex/cyclosome[27] that are also RING dependent. Our data suggest E3 ligases for other Ubls are also likely to use a similar catalytic mechanism[23,25]. The unifying concept is that the E3 activates E2–Ub/Ubl thioester by holding the Ub/Ubl in the folded-back position, extending its C-terminal tail. This is akin to tensioning a spring that would be released by cleavage of the thioester and formation of the isopeptide bond. Although details of the molecular contacts that fold back the Ub/Ubl will vary, it is the position of the C-terminal tail of the Ub/Ubl in the active site groove of the E2 that is central to the process.

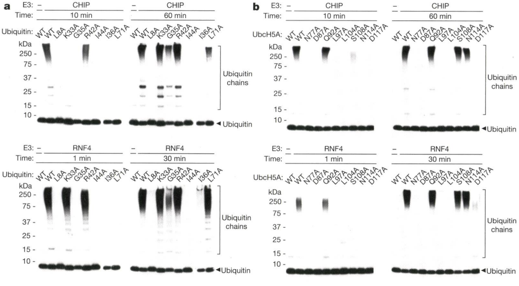

a
E3: – CHIP – CHIP
Time: 10 min 60 min
Ubiquitin: WT WT L8A K33A G35A R42A I44A I36A L71A
kDa
250
75
37
25
15
10
Ubiquitin chains
Ubiquitin
E3: – RNF4 – RNF4
Time: 1 min 30 min
b
E3: – CHIP – CHIP
Time: 10 min 60 min
UbcH5A: WT WT N77A D87A Q92A L97A L104A S108A N114A D117A
E3: – RNF4 – RNF4
Time: 1 min 30 min

**Figure 4 | The same interfaces in E2 and ubiquitin are important for CHIP and RNF4 activity.** **a**, Autoubiquitination activity of CHIP (top panel) and RNF4 (bottom panel) with ubiquitin mutants. Western blots probed with anti-ubiquitin antibody are shown. Longer exposure is shown for I36A and L71A ubiquitin, as binding of the antibody is affected by these mutations. **b**, Autoubiquitination activity as in **a**, but with UbcH5A mutants.

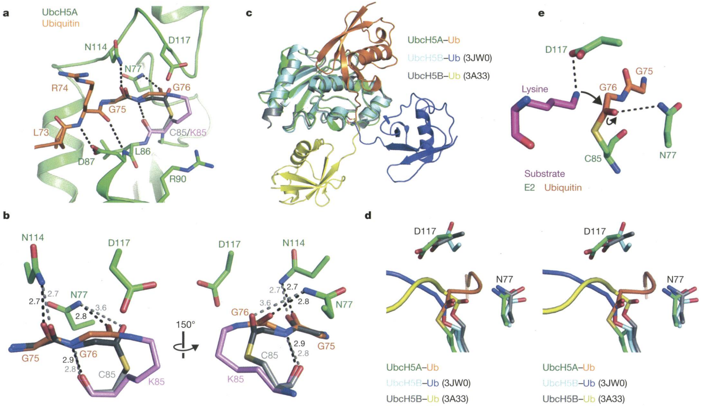


**Figure 5 | E3-mediated structural changes associated with the catalytically primed form of UbcH5A–Ub. a**, Model of UbcH5A–Ub thioester (grey) compared with isopeptide-linked UbcH5A(C85K)–Ub (K85 is violet). **b**, Comparison of modelled thioester with isopeptide linkage. Hydrogen bonds are black (isopeptide) or grey (modelled thioester) dashes, with distances shown in Å. **c**, Comparison of the position of ubiquitin relative to E2 in the UbcH5A–Ub–RING complex reported here with the UbcH5B–Ub–HECT(NEDD4L) complex (Protein Data Bank accession 3JW0)[19], and UbcH5B–Ub oxyester (Protein Data Bank accession 3A33)[18]. **d**, RING-mediated remodelling of the UbcH5A active site. The position of the C terminus of ubiquitin linked to the active site cysteine/serine of the E2 is shown relative to residues N77 and D117 in the three structures shown in **c**. **e**, Model for nucleophilic attack by substrate lysine (pink) on the E2–Ub thioester bond, based on the SUMO–RanGAP1–UBC9–RanBP2 structure[25].

## METHODS SUMMARY

Recombinant proteins were expressed in *Escherichia coli* cells and purified by standard methods. For structural analysis of a stable mimic of the UbcH5A–Ub thioester, mutations C85K and S22R[28] were introduced into UbcH5A (UbcH5A(S22R/C85K)). The isopeptide bond-linked UbcH5A(S22R/C85K)–Ub conjugate was prepared by incubating UbcH5A(S22R/C85K) (200 $\mu$M) with His$_6$-tagged ubiquitin (200 $\mu$M) and E1 (1 $\mu$M) at 35 °C for 26 h in a buffer containing 3 mM ATP, 5 mM MgCl$_2$, 50 mM Tris pH 10.0, 150 mM NaCl and 0.8 mM TCEP. The E2–Ub conjugate was purified by Ni$^{2+}$-affinity chromatography. His$_6$-tag was removed using TEV protease and the conjugate was further purified by Ni$^{2+}$-affinity chromatography and gel filtration chromatography. The RNF4 RING–UbcH5A(S22R/C85K)–Ub complex was prepared by mixing the UbcH5A(S22R/C85K)–Ub with a linear fusion of two RNF4 RING domains in a 2:1 molar ratio. Crystals grew from a 1:1 sitting-drop with a reservoir solution containing 18% (w/v) PEG 3,000, 0.1 M Tris (pH 7.2), and 0.2 M calcium acetate. The structure was solved by molecular replacement to a resolution of 2.2 Å using in house X-rays. A single-turnover substrate ubiquitination assay for RNF4 has been described previously[3].

**Full Methods** and any associated references are available in the online version of the paper.

**Received 19 March; accepted 10 July 2012.**
**Published online 29 July 2012.**

1. Kravtsova-Ivantsiv, Y. & Ciechanover, A. Non-canonical ubiquitin-based signals for proteasomal degradation. *J. Cell Sci.* **125,** 539–548 (2012).
2. Budhidarmo, R., Nakatani, Y. & Day, C. L. RINGs hold the key to ubiquitin transfer. *Trends Biochem. Sci.* **37,** 58–65 (2012).
3. Plechanovová, A. *et al.* Mechanism of ubiquitylation by dimeric RING ligase RNF4. *Nature Struct. Mol. Biol.* **18,** 1052–1059 (2011).
4. Galanty, Y., Belotserkovskaya, R., Coates, J. & Jackson, S. P. RNF4, a SUMO-targeted ubiquitin E3 ligase, promotes DNA double-strand break repair. *Genes Dev.* **26,** 1179–1195 (2012).
5. Luo, K., Zhang, H., Wang, L., Yuan, J. & Lou, Z. Sumoylation of MDC1 is important for proper DNA damage response. *EMBO J.* **31,** 3008–3019 (2012).
6. Yin, Y. *et al.* SUMO-targeted ubiquitin E3 ligase RNF4 is required for the response of human cells to DNA damage. *Genes Dev.* **26,** 1196–1208 (2012).
7. Lallemand-Breitenbach, V. *et al.* Arsenic degrades PML or PML-RARα through a SUMO-triggered RNF4/ubiquitin-mediated pathway. *Nature Cell Biol.* **10,** 547–555 (2008).
8. Tatham, M. H. *et al.* RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. *Nature Cell Biol.* **10,** 538–546 (2008).
9. Liew, C. W., Sun, H., Hunter, T. & Day, C. L. RING domain dimerization is essential for RNF4 function. *Biochem. J.* **431,** 23–29 (2010).
10. Mace, P. D. *et al.* Structures of the cIAP2 RING domain reveal conformational changes associated with ubiquitin-conjugating enzyme (E2) recruitment. *J. Biol. Chem.* **283,** 31633–31640 (2008).
11. Bentley, M. L. *et al.* Recognition of UbcH5c and the nucleosome by the Bmi1/Ring1b ubiquitin ligase complex. *EMBO J.* **30,** 3285–3297 (2011).
12. Bosanac, I. *et al.* Modulation of K11-linkage formation by variable loop residues within UbcH5A. *J. Mol. Biol.* **408,** 420–431 (2011).
13. Bosanac, I. *et al.* Ubiquitin binding to A20 ZnF4 is required for modulation of NF-kappaB signaling. *Mol. Cell* **40,** 548–557 (2010).
14. Zhang, L. *et al.* The IDOL-UBE2D complex mediates sterol-dependent degradation of the LDL receptor. *Genes Dev.* **25,** 1262–1274 (2011).
15. Hamilton, K. S. *et al.* Structure of a conjugating enzyme-ubiquitin thiolester intermediate reveals a novel role for the ubiquitin tail. *Structure* **9,** 897–904 (2001).
16. Pruneda, J. N., Stoll, K. E., Bolton, L. J., Brzovic, P. S. & Klevit, R. E. Ubiquitin in motion: structural studies of the ubiquitin-conjugating enzyme~ubiquitin conjugate. *Biochemistry* **50,** 1624–1633 (2011).
17. Wickliffe, K. E., Lorenz, S., Wemmer, D. E., Kuriyan, J. & Rape, M. The mechanism of linkage-specific ubiquitin chain elongation by a single-subunit E2. *Cell* **144,** 769–781 (2011).
18. Sakata, E. *et al.* Crystal structure of UbcH5b~ubiquitin intermediate: insight into the formation of the self-assembled E2~Ub conjugates. *Structure* **18,** 138–147 (2010).
19. Kamadurai, H. B. *et al.* Insights into ubiquitin transfer cascades from a structure of a UbcH5B~ubiquitin-HECT(NEDD4L) complex. *Mol. Cell* **36,** 1095–1102 (2009).
20. Wu, P. Y. *et al.* A conserved catalytic residue in the ubiquitin-conjugating enzyme family. *EMBO J.* **22,** 5241–5250 (2003).
21. Yunus, A. A. & Lima, C. D. Lysine activation and functional analysis of E2-mediated conjugation in the SUMO pathway. *Nature Struct. Mol. Biol.* **13,** 491–499 (2006).

22. Yunus, A. A. & Lima, C. D. Structure of the Siz/PIAS SUMO E3 ligase Siz1 and determinants required for SUMO modification of PCNA. *Mol. Cell* **35,** 669–682 (2009).
23. Reverter, D. & Lima, C. D. Insights into E3 ligase activity revealed by a SUMO-RanGAP1-Ubc9-Nup358 complex. *Nature* **435,** 687–692 (2005).
24. Saha, A., Lewis, S., Kleiger, G., Kuhlman, B. & Deshaies, R. J. Essential role for ubiquitin-ubiquitin-conjugating enzyme interaction in ubiquitin discharge from Cdc34 to substrate. *Mol. Cell* **42,** 75–83 (2011).
25. Gareau, J. R., Reverter, D. & Lima, C. D. Determinants of small ubiquitin-like modifier 1 (SUMO1) protein specificity, E3 ligase, and SUMO-RanGAP1 binding activities of nucleoporin RanBP2. *J. Biol. Chem.* **287,** 4740–4751 (2012).
26. Calabrese, M. F. *et al.* A RING E3-substrate complex poised for ubiquitin-like protein transfer: structural insights into cullin-RING ligases. *Nature Struct. Mol. Biol.* **18,** 947–949 (2011).
27. Schreiber, A. *et al.* Structural basis for the subunit assembly of the anaphase-promoting complex. *Nature* **470,** 227–232 (2011).
28. Brzovic, P. S., Lissounov, A., Christensen, D. E., Hoyt, D. W. & Klevit, R. E. A UbcH5/ubiquitin noncovalent complex is required for processive BRCA1-directed ubiquitination. *Mol. Cell* **21,** 873–880 (2006).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** We thank M. Alphey and E. Branigan for assistance with data collection. His–UBE1 was a gift from the Division of Signal Transduction Therapy, University of Dundee. CHIP was a gift from A. Knebel and P. Cohen. A.P. was funded by the Wellcome Trust. This work was supported by a grant to R.T.H. from Cancer Research UK. Structural biology was supported by Scottish Funding Council (ref SULSA) and Wellcome Trust (program grant JHN).

**Author Contributions** A.P. cloned, expressed and purified proteins, carried out structural analysis, conducted biochemical experiments and interpreted the data. E.G.J. purified recombinant proteins and carried out biochemical analysis. M.H.T. carried out mass spectrometry analysis. J.H.N. contributed to structural analysis and data analysis. A.P., J.H.N. and R.T.H. wrote the paper. R.T.H. conceived the project and contributed to data analysis.

**Author Information** Coordinates and structure factors of the RNF4 RING–UbcH5A(S22R/C85K)–Ub complex were deposited in the Protein Data Bank under accession code 4AP4. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.T.H. (r.t.hay@dundee.ac.uk).

doi:10.1038/nature11407

# Observation of interstellar lithium in the low-metallicity Small Magellanic Cloud

J. Christopher Howk[1], Nicolas Lehner[1], Brian D. Fields[2,3] & Grant J. Mathews[1]

**The primordial abundances of light elements produced in the standard theory of Big Bang nucleosynthesis (BBN) depend only on the cosmic ratio of baryons to photons, a quantity inferred from observations of the microwave background[1]. The predicted[2–4] primordial $^7$Li abundance is four times that measured in the atmospheres of Galactic halo stars[5–7]. This discrepancy could be caused by modification of surface lithium abundances during the stars' lifetimes[8] or by physics beyond the Standard Model that affects early nucleosynthesis[9,10]. The lithium abundance of low-metallicity gas provides an alternative constraint on the primordial abundance and cosmic evolution of lithium[11] that is not susceptible to the *in situ* modifications that may affect stellar atmospheres. Here we report observations of interstellar $^7$Li in the low-metallicity gas of the Small Magellanic Cloud, a nearby galaxy with a quarter the Sun's metallicity. The present-day $^7$Li abundance of the Small Magellanic Cloud is nearly equal to the BBN predictions, severely constraining the amount of possible subsequent enrichment of the gas by stellar and cosmic-ray nucleosynthesis. Our measurements can be reconciled with standard BBN with an extremely fine-tuned depletion of stellar Li with metallicity. They are also consistent with non-standard BBN.**

We obtained high-resolution spectra ($R \approx 70{,}000$) of the star Sk 143 (AzV 456), an O-type supergiant star in the Small Magellanic Cloud (SMC), using the Ultraviolet and Visual Echelle Spectrograph (UVES)[12] on the 8.2-m Very Large Telescope (VLT); observational details are given in the Supplementary Information. The sight line to this star was chosen for observation because it shows significant absorption from neutral atoms and molecules[13–15] and a weak interstellar radiation field[14], all of which favour the presence of neutral lithium (Li I). Li I absorption is clearly detected along this sight line (Fig. 1).

The derivation of the total Li/H abundance in the interstellar medium (ISM) requires large corrections for ionization, given the column density of Li, $N(\mathrm{Li}) \approx N(\mathrm{Li\,II}) \gg N(\mathrm{Li\,I})$, and for the incorporation of Li into interstellar dust grains[16]. Our first approach to these corrections uses observations of adjacent ionization states of other metals, in this case Ca and Fe, to estimate the amount of unseen gas-phase lithium. Assuming ionization balance and only atomic processes, we have the ratios $N(\mathrm{Li\,II})/N(\mathrm{Li\,I}) \propto N(\mathrm{Ca\,II})/N(\mathrm{Ca\,I})$ or $N(\mathrm{Li\,II})/N(\mathrm{Li\,I}) \propto N(\mathrm{Fe\,II})/N(\mathrm{Fe\,I})$, where the constants of proportionality involve the ratios of ionization rates and recombination coefficients for the elements in question[16,17]. The ratio of $^7$Li I to total hydrogen in the SMC is $\log[N(^7\mathrm{Li\,I})/N(\mathrm{H})] = -11.17 \pm 0.04$ (all uncertainties are $1\sigma$ unless noted), where $N(\mathrm{H}) \equiv N(\mathrm{H\,I}) + 2N(\mathrm{H_2})$. Applying ionization corrections derived from Ca and Fe yields logarithmic abundances $A(^7\mathrm{Li}) \equiv \log[N(^7\mathrm{Li})/N(\mathrm{H})] + 12 = 2.79 \pm 0.11$ (from Ca) and $3.01 \pm 0.12$ (from Fe). These calculations do not include more complicated (and uncertain) effects such as grain-assisted recombination[17,18], nor do they correct for dust depletion.

Our second approach uses the observation[16] that $N(^7\mathrm{Li\,I})/N(\mathrm{K\,I})$ along sight lines through the Milky Way is nearly constant (with new determinations giving consistent results[19,20]). When a differential ionization correction is applied, $^7$Li/K in the Milky Way ISM is consistent with the Solar System ratio. Thus, $^7$Li and K appear to have very similar ionization and dust depletion behaviours, and $^7$Li I/K I gives a good measure of the total (gas + dust phase) $^7$Li/K (refs 16, 19 and 20). We measure $\log[N(^7\mathrm{Li\,I})/N(\mathrm{K\,I})] = -2.27 \pm 0.03$ in the SMC, in agreement with the Galactic relationship[19,20]. Applying an ionization correction of $+0.54 \pm 0.08$ dex (refs 16 and 17) gives $\log[N(^7\mathrm{Li})/N(\mathrm{K})] = -1.78 \pm 0.09$. With the Solar System ratio $\log(^7\mathrm{Li}/\mathrm{K})_\odot = -1.82 \pm 0.05$ derived from meteorites[21], we find $[^7\mathrm{Li}/\mathrm{K}]_{\mathrm{SMC}} \equiv \log[N(^7\mathrm{Li})/N(\mathrm{K})] - \log(^7\mathrm{Li}/\mathrm{K})_\odot$. The ratio of $^7$Li to metal nuclei in the SMC is consistent with that found in the Solar System and the Milky Way ISM[16]: $(^7\mathrm{Li}/\mathrm{K})_{\mathrm{SMC}} \approx (^7\mathrm{Li}/\mathrm{K})_\odot$.

Although the ionization and depletion characteristics of S I are not as well tied to those of Li I (ref. 17), a similar approach using S I yields $[^7\mathrm{Li}/\mathrm{S}]_{\mathrm{SMC}} = -0.26 \pm 0.11$. The sub-solar ratio is consistent with a modest (0.3 dex) depletion of Li and K onto dust in the ISM[19] relative to S.

We estimate $A(^7\mathrm{Li})$ by scaling $^7$Li/K to Li/H: $A(^7\mathrm{Li})_{\mathrm{SMC}} = A(^7\mathrm{Li})_\odot + [\mathrm{Fe/H}]_{\mathrm{SMC}} + [\mathrm{K/Fe}]_{\mathrm{SMC}} + [^7\mathrm{Li}/\mathrm{K}]_{\mathrm{SMC}}$. We adopt $[^7\mathrm{Li}/\mathrm{K}]_{\mathrm{SMC}}$ from above, the meteoritic $A(^7\mathrm{Li})_\odot = 3.23 \pm 0.05$ (ref. 21), with a mean present-day SMC metallicity $[\mathrm{Fe/H}]_{\mathrm{SMC}} = -0.59 \pm 0.06$ and an SMC K/Fe abundance $[\mathrm{K/Fe}]_{\mathrm{SMC}} \equiv +0.00 \pm 0.10$ (these last two are discussed in the Supplementary Information). This yields $A(^7\mathrm{Li})_{\mathrm{SMC}} = 2.68 \pm 0.16$. Similarly scaling the $^7$Li/S result gives $2.38 \pm 0.17$.

Most previous observational constraints on the primordial Li abundance have relied on measurements of atmospheric abundances in low-metallicity Galactic stars. Our detection of interstellar lithium beyond the Milky Way opens a new window on the lithium problem. Although there are significant uncertainties associated with ionization and dust effects, as demonstrated by the significant spread in $A(^7\mathrm{Li})_{\mathrm{SMC}}$ values, these are largely independent of the uncertainties that might affect stellar measurements of the primordial lithium abundance. Our recommended absolute abundance is $A(^7\mathrm{Li})_{\mathrm{SMC}} = 2.68 \pm 0.16$, or $(^7\mathrm{Li}/\mathrm{H})_{\mathrm{SMC}} = (4.8 \pm 1.8) \times 10^{-10}$, derived from $^7$Li/K. This is compared to stellar $^7$Li abundances[6,22] at different metallicities in Fig. 2. Our best estimate overlaps the prediction from standard BBN using the baryonic density deduced from the five-year Wilkinson Microwave Anisotropy Probe (WMAP) data[1], $A(^7\mathrm{Li}) = 2.72 \pm 0.06$ (95% confidence level; ref. 3), although this leaves little room for the post-BBN chemical evolution[23,24], that is, the contribution of freshly synthesized Li to the ISM by stellar and cosmic ray nucleosynthesis (see representative models[23] in Fig. 2). Our estimate of $A(^7\mathrm{Li})_{\mathrm{SMC}}$ is also consistent with the upper envelope of Li abundances in Milky Way thin-disk stars (Fig. 2)[22].

However, given the uncertainties in scaling to $A(^7\mathrm{Li})_{\mathrm{SMC}}$, the stronger result is our measurement of $[^7\mathrm{Li}/\mathrm{K}]_{\mathrm{SMC}} = +0.04 \pm 0.10$. We compare $[^7\mathrm{Li}/\mathrm{K}]_{\mathrm{SMC}}$ with measurements[6,21,22] of $[^7\mathrm{Li}/\mathrm{Fe}]$ and chemical evolution models[23] in Fig. 3. The stars show a rapid decrease in $[^7\mathrm{Li}/\mathrm{Fe}]$ with increasing metallicity until $[\mathrm{Fe/H}] \approx -1$, at which point the Li abundance increases roughly in lockstep with Fe such that disk stars have a

---

[1]Department of Physics, Center for Astrophysics, University of Notre Dame, Notre Dame, Indiana 46556, USA. [2]Department of Astronomy, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. [3]Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA.

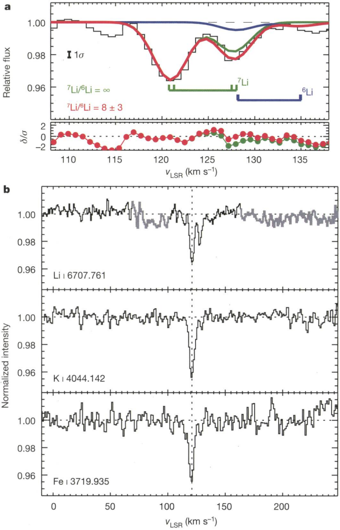


**Figure 1 | Interstellar absorption by several neutral species seen towards the star Sk 143.** Normalized interstellar absorption profiles from UVES plotted versus the Local Standard of Rest velocity, $v_{LSR}$, and profile fit of the Li I absorption. The empirically determined signal-to-noise ratio is about 275 per pixel (5 pixels per resolution element) for the Li I observations. The full set of optical and ultraviolet absorption profiles seen towards this star and the column densities measured from these are given in the Supplementary Information. **b**, The profiles of Li I, K I, and Fe I; the SMC cloud bearing Li I at $v_{LSR} \approx +121$ km s$^{-1}$ is marked with the dashed line. The thicker grey regions near Li I are possibly contaminated by diffuse interstellar bands or residual fringing, which may extend into the region containing Li absorption. The effects on the $^7$Li I columns are within the quoted uncertainties. The Li I absorption is composed of (hyper)fine structure components of both $^7$Li I and $^6$Li I (shown, respectively, by the green and blue ticks in the top panel of **a**). The strong line of $^7$Li I is detected with approximately 16σ significance in the ISM of the SMC. A model fit to the Li I absorption complex is shown in **a** (see Supplementary Information), with the difference between the data and the fit, δ, shown immediately below (normalized to the local error array). The free parameters for the fit are the polynomial coefficients for the stellar continuum, the central velocity, Doppler parameter (*b*-value), and column densities of $^7$Li I and $^6$Li I for the interstellar cloud. The red curve shows the best-fitting model including both $^7$Li I and $^6$Li I, which are shown in green and blue, respectively. The best-fit isotopic ratio is $N(^6\mathrm{Li\,I})/N(^7\mathrm{Li\,I}) = 0.13 \pm 0.05$ (68% confidence limit), consistent with the presence of $^6$Li along the sight line, although below the 3σ detection threshold.

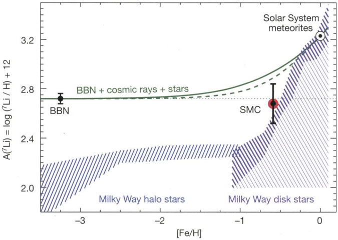


**Figure 2 | Estimates of the lithium abundance in the SMC interstellar medium and in other environments.** Our best estimate for the interstellar (gas + dust phase) abundance of $A(^7\mathrm{Li})$ in the SMC (red circle) is derived from the $^7$Li I/K I ratio. The present day metallicity of the SMC from early-type stars is [Fe/H] = −0.59 ± 0.06. (All uncertainties are 1σ.) The point marked BBN and the dotted horizontal line show the primordial abundance predicted by standard BBN[3]. The green curves show recent models[23] for post-BBN $^7$Li nucleosynthesis due to cosmic rays and stars. By adjusting the yields from low-mass stars, the models are forced to match the Solar System meteoritic abundance[21] (see Supplementary Information). The solid and dashed lines correspond to models A and B[23], which include (A) or do not include (B) a presumed contribution to $^7$Li from core-collapse supernovae. The blue hatched area shows the range of abundances derived for Population II stars in the Galactic halo[6], with the 'Spite plateau' in this sample at $A(^7\mathrm{Li})_{\mathrm{Pop\,II}} \approx 2.10 \pm 0.10$ (ref. 6). The violet hatched area shows the range of measurements seen in Galactic thin-disk stars, and the thicker violet lines denote the six most Li-rich stars in a series of eight metallicity bins[22]. The selection of thin-disk stars includes objects over a range of masses and temperatures, including stars that are expected to have destroyed a fair fraction of their Li. Thus, the upper envelope of the distribution represents the best estimate of the intrinsic ISM Li abundance at the epoch of formation for those stars, and the thicker hatched area for the thin-disk sample is most appropriate for comparison with the SMC value. The most Li-rich stars in the Milky Way thin disk[22] within 0.1 dex of the SMC metallicity give $A(^7\mathrm{Li})_{\mathrm{Milky\,Way}} = 2.54 \pm 0.05$, consistent with our estimate of $A(^7\mathrm{Li})_{\mathrm{SMC}} = 2.68 \pm 0.16$.

nearly constant [$^7$Li/Fe] ratio, similar to that found in the Solar System. Our measurement of the present-day $^7$Li-to-metal ratio in the SMC is in agreement with the nearly constant values found in the atmospheres of Milky Way disk stars ($-1 \lesssim$ [Fe/H] $\lesssim 0$), most of which formed over 4 billion years ago, with the Solar System and the modern-day Milky Way ISM[16].

Both the thin-disk stars and our SMC measurements are below standard BBN predictions with reasonable assumptions about post-BBN production, although it is often assumed these stars have had significant depletion of their surface Li abundance[23]. Taken at face value, the consistency of our SMC measurement with the [$^7$Li/Fe] for those stars calls this assumption into question. Although the models in Figs 2 and 3 are imprecise given the uncertain Li yields from stellar sources, they illustrate the tension between standard BBN predictions and our measurements if there is any post-BBN Li production. This tension can be relieved if a metallicity-dependent depletion of Li in stellar atmospheres is fine-tuned in such a way that it is very strong below [Fe/H] ≈ [Fe/H]$_{\mathrm{SMC}}$ = −0.6 (to create the Spite plateau and avoid overproducing Li in the SMC ISM) and negligible at or above the SMC metallicity, thus conspiring to create a constant [$^7$Li/Fe] ratio above [Fe/H] ≈ −1. Alternatively, non-standard BBN scenarios can be invoked to allow for a lower primordial Li abundance[4,25].

If non-standard Li production occurs in the BBN epoch, many such models predict excess $^6$Li compared with the standard BBN. The only known source of post-Big Bang $^6$Li is production via cosmic ray interactions with ISM particles. Excess $^6$Li at the metallicity of the SMC would support non-standard production mechanisms, either in the BBN epoch[10] or through the interaction of pregalactic cosmic rays with intergalactic helium[26]. Measurements of $^6$Li in stellar atmospheres are extremely difficult because the stellar line broadening is well in excess

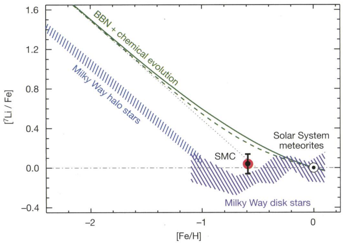


**Figure 3 | Estimates of Li/Fe in the SMC interstellar medium and in several different environments.** The SMC value is derived from the $^7$Li I/K I ratio. At low metallicities ($[\mathrm{Fe/H}] \lesssim -1$), stellar measurements[6] trace the build-up of Fe with a constant Li abundance along the Spite plateau. At higher metallicities, disk star abundances[22] show a turnover to roughly constant [$^7$Li/Fe] at values consistent with the Solar System meteoritic value[21] (shown as the dash-dotted black line at $[^7\mathrm{Li/Fe}] = 0$). Our SMC estimate is consistent with the Solar System and disk star abundances in this region of relatively constant $^7$Li/Fe abundances, with $[^7\mathrm{Li/Fe}]_{\mathrm{SMC}} = +0.04 \pm 0.14$ for $[\mathrm{K/Fe}]_{\mathrm{SMC}} = 0.0 \pm 0.10$ (Supplementary Information). The most Li-rich disk stars within 0.1 dex of the SMC metallicity have a mean $[^7\mathrm{Li/Fe}] = -0.13 \pm 0.05$. (All uncertainties are $1\sigma$.) The green curves show the chemical evolution models[23] as in Fig. 2, whereas the dotted line shows the behaviour of [$^7$Li/Fe] for the standard BBN primordial abundance with no subsequent evolution of $^7$Li. The relative uniformity of the stellar $^7$Li/Fe abundances at $[\mathrm{Fe/H}] \gtrsim -1$ could be caused by a delicate balance of Li and Fe production and metallicity-dependent depletion of the surface Li abundances (not ruled out given the changes in mean age and mass potentially present in the sample[22]). However, the agreement of the [$^7$Li/Fe] ratio seen in these old stars (ages exceeding 4 billion years[22]) and in the present-day interstellar medium of the SMC suggests little change in the stellar abundances for metallicities $[\mathrm{Fe/H}] \approx -0.6$ up to the solar metallicity. To bring the stellar and SMC interstellar abundances into agreement with standard BBN predictions requires a delayed injection of significant $^7$Li from stellar production mechanisms as well as vigorous depletion of stellar surface $^7$Li abundances at metallicities just below that of the SMC.

of the isotope shift. However, the $^7$Li I doublet is well separated in our data owing to the very low broadening in the cool ISM probed by Li I absorption. Our best fit to the SMC Li I absorption gives $(^6\mathrm{Li}/^7\mathrm{Li})_{\mathrm{SMC}} = 0.13 \pm 0.05$ (see Supplementary Information and Fig. 1), giving a formal limit to the isotopic ratio in the SMC of $(^6\mathrm{Li}/^7\mathrm{Li})_{\mathrm{SMC}} < 0.28$ $(3\sigma)$. With higher signal-to-noise ratios and resolution it should be possible to lower the limits for the interstellar isotope ratio in the SMC to provide constraints on non-standard BBN models. This approach has the advantage that ionization and dust-depletion effects are not important for comparing the two isotopes of Li (ref. 27), making $^6$Li/$^7$Li a powerful diagnostic of nucleosynthesis and non-standard evolution of Li abundances.

**Received 1 June; accepted 16 July 2012.**

1. Dunkley, J. *et al.* Five-year Wilkinson Microwave Anisotropy Probe observations: likelihoods and parameters from the WMAP data. *Astrophys. J.* **180** (Suppl.), 306–329 (2009).
2. Steigman, G. Primordial nucleosynthesis in the precision cosmology era. *Ann. Rev. Nuclear Particle Sci.* **57,** 463–491 (2007).
3. Cyburt, R. H., Fields, B. D. & Olive, K. A. An update on the big bang nucleosynthesis prediction for $^7$Li: the problem worsens. *J. Cosmol. Astro-Particle Phys.* **11,** 012 (2008).
4. Fields, B. D. The primordial lithium problem. *Ann. Rev. Nuclear Particle Sci.* **61,** 47–68 (2011).
5. Spite, M. & Spite, F. Lithium abundance at the formation of the Galaxy. *Nature* **297,** 483–485 (1982).
6. Sbordone, L. *et al.* The metal-poor end of the Spite plateau. 1: Stellar parameters, metallicities and lithium abundances. *Astron. Astrophys.* **522,** A26 (2010).
7. Meléndez, J., Casagrande, L., Ramírez, I., Asplund, M. & Schuster, W. J. Observational evidence for a broken Li Spite plateau and mass-dependent Li depletion. *Astron. Astrophys.* **515,** L3–L7 (2010).
8. Korn, A. J. *et al.* A probable stellar solution to the cosmological lithium discrepancy. *Nature* **442,** 657–659 (2006).
9. Jedamzik, K. Did something decay, evaporate, or annihilate during big bang nucleosynthesis? *Phys. Rev. D* **70,** 063524 (2004).
10. Pospelov, M. & Pradler, J. Big Bang nucleosynthesis as a probe of new physics. *Ann. Rev. Nuclear Particle Sci.* **60,** 539–568 (2010).
11. Prodanović, T. & Fields, B. D. Probing primordial and pre-galactic lithium with high-velocity clouds. *Astrophys. J.* **616,** L115–L118 (2004).
12. Dekker, H., D'Odorico, S., Kaufer, A., Delabre, B. & Kotzlowski, H. Design, construction, and performance of UVES, the echelle spectrograph for the UT2 Kueyen Telescope at the ESO Paranal Observatory. *Proc. SPIE* **4008,** 534–545 (2000).
13. Cox, N. L. J. *et al.* Interstellar gas, dust and diffuse bands in the SMC. *Astron. Astrophys.* **470,** 941–955 (2007).
14. Welty, D. E., Federman, S. R., Gredel, R., Thorburn, J. A. & Lambert, D. L. VLT UVES observations of interstellar molecules and diffuse bands in the Magellanic clouds. *Astrophys. J.* **165** (Suppl.), 138–172 (2006).
15. Cartledge, S. I. B. *et al.* FUSE measurements of far-ultraviolet extinction. II. Magellanic cloud sight lines. *Astrophys. J.* **630,** 355–367 (2005).
16. Steigman, G. Cosmic lithium: going up or coming down? *Astrophys. J.* **457,** 737–742 (1996).
17. Welty, D. E., Hobbs, L. M. & Morton, D. C. High-resolution observations of interstellar Ca I absorption-implications for depletions and electron densities in diffuse clouds. *Astrophys. J.* **147** (Suppl.), 61–96 (2003).
18. Weingartner, J. C. & Draine, B. T. Electron-ion recombination on grains and polycyclic aromatic hydrocarbons. *Astrophys. J.* **563,** 842–852 (2001).
19. Knauth, D. C., Federman, S. R. & Lambert, D. L. An ultra-high-resolution survey of the interstellar $^6$Li/$^7$Li isotope ratio in the solar neighborhood. *Astrophys. J.* **586,** 268–285 (2003).
20. Welty, D. E. & Hobbs, L. M. A. High-resolution survey of interstellar K I absorption. *Astrophys. J.* **133** (Suppl.), 345–393 (2001).
21. Asplund, M., Grevesse, N., Sauval, A. J. & Scott, P. The chemical composition of the sun. *Annu. Rev. Astron. Astrophys.* **47,** 481–522 (2009).
22. Lambert, D. L. & Reddy, B. E. Lithium abundances of the local thin disc stars. *Mon. Not. R. Astron. Soc.* **349,** 757–767 (2004).
23. Prantzos, N. Production and evolution of Li, Be and B isotopes in the Galaxy. *Astron. Astrophys.* **542,** A67 (2012).
24. Romano, D., Tosi, M., Matteucci, F. & Chiappini, C. Light element evolution resulting from WMAP data. *Mon. Not. R. Astron. Soc.* **346,** 295–303 (2003).
25. Iocco, F., Mangano, G., Miele, G., Pisanti, O. & Serpico, P. D. Primordial nucleosynthesis: from precision cosmology to fundamental physics. *Phys. Rep.* **472,** 1–76 (2009).
26. Suzuki, T. K. & Inoue, S. Cosmic-ray production of $^6$Li by structure formation shocks in the early Milky Way: a fossil record of dissipative processes during galaxy formation. *Astrophys. J.* **573,** 168–173 (2002).
27. Kawanomoto, S. *et al.* The new detections of $^7$Li/$^6$Li isotopic ratio in the interstellar media. *Astrophys. J.* **701,** 1506–1518 (2009).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank the European Southern Observatory for granting us time for this project as part of proposal 382.B-0556. We also thank A. Fox and H. Sana for discussions about the UVES data and A. Korn, P. Molaro, T. Prodanovic, D. Romano, and D. Welty for input on the project that improved the paper.

**Author Contributions** All authors participated in the interpretation and commented on the manuscript. J.C.H. led the project and was responsible for the text of the paper.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.C.H. (jhowk@nd.edu).

LETTER

doi:10.1038/nature11441

# No meridional plasma flow in the heliosheath transition region

Robert B. Decker[1], Stamatios M. Krimigis[1,2], Edmond C. Roelof[1] & Matthew E. Hill[1]

**Over a two-year period, Voyager 1 observed a gradual slowing-down of radial plasma flow in the heliosheath to near-zero velocity[1] after April 2010 at a distance of 113.5 astronomical units from the Sun (1 astronomical unit equals $1.5 \times 10^8$ kilometres). Voyager 1 was then about 20 astronomical units beyond the shock that terminates the free expansion of the solar wind and was immersed in the heated non-thermal plasma region called the heliosheath. The expectation from contemporary simulations[2,3] was that the heliosheath plasma would be deflected from radial flow to meridional flow (in solar heliospheric coordinates), which at Voyager 1 would lie mainly on the (locally spherical) surface called the heliopause. This surface is supposed to separate the heliosheath plasma, which is of solar origin, from the interstellar plasma, which is of local Galactic origin. In 2011, the Voyager project began occasional temporary re-orientations of the spacecraft (totalling about 10–25 hours every 2 months) to re-align the Low-Energy Charged Particle instrument on board Voyager 1 so that it could measure meridional flow. Here we report that, contrary to expectations, these observations yielded a meridional flow velocity of $+3 \pm 11$ km s$^{-1}$, that is, one consistent with zero within statistical uncertainties.**

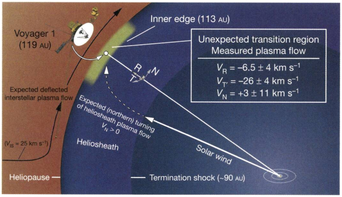

Voyager 1
(119 AU)
Inner edge (113 AU)
Unexpected transition region
Measured plasma flow
$V_R = -6.5 \pm 4$ km s$^{-1}$
$V_{T'} = -26 \pm 4$ km s$^{-1}$
$V_N = +3 \pm 11$ km s$^{-1}$
R
N
T
Expected deflected interstellar plasma flow
Expected (northern) turning of heliosheath plasma flow
$V_N > 0$
($V_{IS}$ = 25 km s$^{-1}$)
Heliosheath
Solar wind
Heliopause
Termination shock (~90 AU)

**Figure 1 | Heliospheric plasma boundaries and flow regions relative to the location of Voyager 1.** The solar wind flows initially radially outwards from the Sun, and in the outer heliosheath its expected meridional deflection becomes parallel to the heliopause. The interstellar flow is from the left in the image; it should be deflected around the heliosheath in the region beyond Voyager 1. Voyager 1 is shown in its own meridional plane (in solar heliospheric coordinates) within the unexpected transition region that it first encountered at a helioradius of ~113 AU (ref. 1; unit vectors of the heliospheric RTN system are indicated). It had been expected that heliosheath plasma flow near the heliopause would have a near-zero radial component and that its meridional component $V_N$ would be a significant fraction of 25 km s$^{-1}$, to be consistent with the distant speed of the local interstellar plasma and its deflection around the heliosheath. However, from the data taken during five rolls of Voyager 1, we have determined that $\langle V_R \rangle = -14 \pm 14$ km s$^{-1}$ and $\langle V_N \rangle = +3 \pm 11$ km s$^{-1}$. We conclude that the roll data taken at Voyager 1 are statistically consistent with $V_N = 0$. Figure adapted from an image online on the Voyager website at the Jet Propulsion Laboratory (http://voyager.jpl.nasa.gov/news/new_region.html).

The discovery by Voyager 1 of the zero radial velocity[1] of heliosheath plasma flow beyond ~113.5 astronomical units (AU) in a previously unsuspected transition region (Fig. 1) led to the suggestion that the initially radial flow in the heliosheath was already being deflected polewards (towards meridional flow), as predicted by typical magnetohydrodynamic models[2,3]. The suggested meridional flow could not be measured by the Low-Energy Charged Particle instrument in the usual orientation of its scanning plane on board Voyager 1, so starting in March 2011 the spacecraft was commanded to rotate about its Earth-pointing axis for about one day every second month to enable the instrument to measure flow speeds in the meridional or R–N plane. (In the RTN coordinate system, R is the radius vector from the Sun, T is in the direction of solar rotation and N completes a right-handed system.) Measurements from five such rolls performed during 2011/066–2012/030 have been analysed so far (date notation is

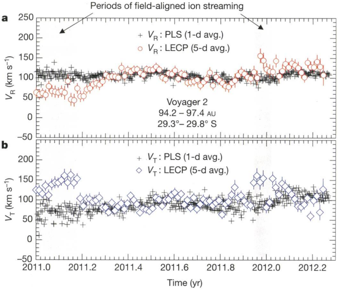


**Figure 2 | Comparison of radial and azimuthal components of heliosheath plasma flow velocity at Voyager 2.** **a**, Crosses, daily averaged values of $V_R$ measured by the plasma instrument (PLS) during 2011/001–2012/035. Circles, 5-d-averaged determinations of $V_R$ using the Fourier fit procedure on 28–43-keV ion angular data from the Low-Energy Charged Particle instrument (LECP). Vertical error bars are Poisson statistical uncertainties ($\pm 1\sigma$) about the mean. During the period shown, Voyager 2 moved from helioradius 94.2 AU to helioradius 97.4 AU and from heliolongitude 29.3° S to heliolongitude 29.8° S. **b**, Crosses, daily-averaged values of $V_T$ measured by the PLS. Diamonds, 5-d-averaged determinations of $V_{T'}$ using the Fourier fit procedure on ion angular data. Vertical error bars are Poisson statistical uncertainties ($2\sigma$) about the mean. There is generally good agreement between the measured solar wind components and those determined from fits to the low-energy ion angular measurements. The two shaded periods show where angular data on ions in several energy channels of the LECP allow us to identify relatively large non-convective anisotropies consistent with $+T$-directed streaming along the average azimuthal orientation of the magnetic field in the heliosheath.

---

[1] Applied Physics Laboratory, The Johns Hopkins University, Laurel, Maryland 20723, USA. [2] Academy of Athens, Athens 11527, Greece.

year/day of year). We report here the absence of a statistically significant persistent N component of flow, with a cumulative average velocity over the five roll periods of $\langle V_N \rangle = +3 \pm 11\,\mathrm{km\,s^{-1}}$. Longer-term averages of the R and T components of flow during 2011/066–2012/030 in the usual instrument orientation yielded a sunward radial velocity of $V_R = -7 \pm 4\,\mathrm{km\,s^{-1}}$ and a persistent negative azimuthal velocity of $V_{T'} = -26 \pm 4\,\mathrm{km\,s^{-1}}$.

We test the null hypothesis for convective flow in the N direction, that is, that the meridional component ($V_N$) of any convective plasma flow within the transition region is statistically consistent with zero. Thus, we make the simplest assumption of convective flow in our most sensitive energy channel (ions with energies of 53–85 keV) and compute the velocity implied by the angular distribution of ion intensities. If that velocity is consistent with zero within our estimated errors, then we conclude that there is no measureable $V_N$.

The Low-Energy Charged Particle telescope samples the anisotropy of the energetic ion intensity in seven positions spaced by 45° in its scan plane; one sector was intentionally blocked. The counting rates $C(\phi_n)$ in the seven usable sector positions ($n = 1, 2, \ldots, 7$) overdetermine the first five coefficients in the Fourier expansion

$$C(\phi) = C_0(1 + A_1 \cos(\phi) + B_1 \sin(\phi) + A_2 \cos(2\phi) + B_2 \sin(2\phi)) \quad (1)$$

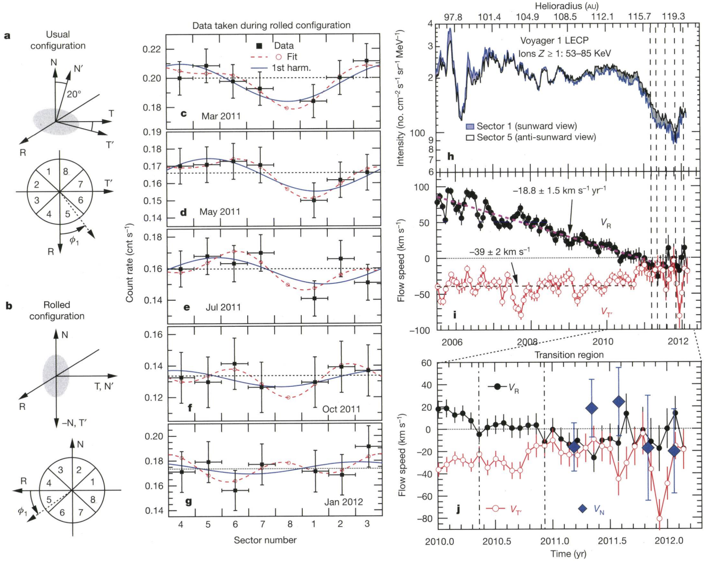


**Figure 3 | Roll measurements and derived components of heliosheath flow velocity.** **a**, Top, orientation of heliospheric (RTN) and instrument-associated (RT′N) axes at the location of Voyager 1 in the usual spacecraft configuration. The LECP scans in the R–T′ plane (shaded grey). The N′–T′ plane is rotated about the R axis by 20° relative to the N–T plane. Bottom, view along the +N′ axis showing sector positions in the R–T′ plane (sector 8 is blocked) and the first-order anisotropy angle $\phi_1$. **b**, Top, orientation of axes in rolled spacecraft configuration. Bottom, view along the +T = +N′ axis showing sector positions in the R–N plane. **c**–**g**, Count rates versus sector in rolled configuration for five roll periods. Solid symbols are roll-averaged count rates of 53–85-keV ions, predominantly protons, based on independent composition measurements. Vertical error bars are Poisson statistical uncertainties ($\pm 1\sigma$) about the mean. Horizontal bars indicate sector angular width. The red curve is a least-squares fit to data of the function in equation (1). The blue curve is the first-harmonic component of the fit. **h**, Intensity of 53–85-keV protons. The blue and black traces are respectively the intensities in sectors 1 and 5 of particles arriving from the sunward (sector 1) and anti-sunward (sector 5) directions in the usual spacecraft configuration. The helioradius of Voyager 1 is given along the top axis. The dashed vertical lines indicate the roll periods in panels **c**–**g**. **i**, 26-d-averaged plasma flow velocity components $V_R$ and $V_{T'}$ (roll periods not included). Vertical error bars are Poisson statistical uncertainties ($2\sigma$) about the mean. For comparison, the average value of the solar wind speed is ~400 km s$^{-1}$. **j**, Expanded view of panel **i** including the five roll determinations of $V_N$ (diamonds). Vertical error bars are Poisson statistical uncertainties ($2\sigma$) about the mean. The first dot–dash vertical line shows the onset, at ~2010/133, of a 208-d (2.05-AU) stretch of zero $V_R$; the second dot–dash vertical line, at ~2010/341, marks the end of the steady zero-$V_R$ flow and the transition to variable and often negative-$V_R$ flow. Anisotropies in the usual configuration after ~2010/341 show non-convective features in sectors 6 and 7, consistent with −T-directed streaming along the average azimuthal orientation of the magnetic field in the heliosheath.

**Table 1 | Voyager 1 roll periods, fit coefficients and heliosheath plasma flow velocities**

| Roll period | No. rolls | $C_0$ ($\times10^{-1}$ cnt s$^{-1}$) | $A_1$ ($\times10^{-2}$) | $B_1$ ($\times10^{-2}$) | $\xi_1$ ($\times10^{-2}$)‡ | $\phi_1$ (°)‡ | $\gamma$ | $V_R$ (km s$^{-1}$) | $V_N$ (km s$^{-1}$) |
|---|---|---|---|---|---|---|---|---|---|
| 1: 2011/066–073 (March 2011) | 6 (21 h)* | 1.97 ± 0.05 (890)† | −6.54 ± 3.78 | −0.20 ± 2.94 | 6.54 ± 3.77 | 159 ± 26 | 1.55 | −26.0 ± 25.8 (28,858 s)§ | −16.3 ± 21.7 |
| 2: 2011/121–131 (May 2011) | 7 (25 h) | 1.65 ± 0.04 (819) | −3.81 ± 4.03 | −4.37 ± 3.13 | 5.80 ± 3.54 | 206 ± 36 | 1.52 | −19.8 ± 27.8 (34,474 s) | +18.4 ± 23.3 |
| 3: 2011/207–217 (July 2011) | 6 (20 h) | 1.58 ± 0.05 (631) | −2.53 ± 4.65 | −4.67 ± 3.62 | 5.32 ± 3.88 | 219 ± 48 | 1.48 | −12.7 ± 32.6 (27,821 s) | +24.2 ± 27.4 |
| 4: 2011/302–307 (October 2011) | 6 (10 h) | 1.32 ± 0.07 (254) | −3.88 ± 7.87 | 0.95 ± 6.13 | 4.00 ± 7.78 | 144 ± 90 | 1.44 | −6.5 ± 56.1 (13,306 s) | −17.3 ± 47.1 |
| 5: 2012/016–030 (January 2012) | 7 (10 h) | 1.74 ± 0.07 (343) | −1.24 ± 6.18 | 2.42 ± 4.73 | 2.73 ± 5.06 | 94 ± 124 | 1.40 | +15.5 ± 44.0 (13,824 s) | −20.2 ± 37.0 |
| Time-weighted average | — | — | — | — | — | — | — | −14.0 ± 13.6‖ | +2.8 ± 11.4 |

* Number of hours of data taken during rolled configuration that were used in ($V_R$, $V_N$) analysis.
† Mean number of counts per sector for the seven active sectors of the Low-Energy Charged Particle instrument.
‡ First-order anisotropy amplitude $\xi_1 = (A_1{}^2 + B_1{}^2)^{1/2}$ and associated azimuth angle $\phi_1 = \tan^{-1}(B_1/A_1)$ (Fig. 3a).
§ Total data accumulation time of data used in flow velocity determination; used to perform weighted average in row 6.
‖ Mean values of $V_R$ and $V_T$ determined from data taken during 2011/066–2012/030 are $V_R = -6.5 \pm 4.1$ km s$^{-1}$ and $V_{T'} = -25.8 \pm 3.8$ km s$^{-1}$.

A least-squares solution yields the amplitudes and phases of the first two harmonics. We assume that ions have an isotropic intensity $j \propto E^{-\gamma}$ in a frame moving with the heliosheath flow. The instrument measures the spectral slope ($\gamma$) in adjacent energy ($E$) channels. The well-known theory of the Compton–Getting effect[4] relates the components of the convective flow ($V_R$, $V_N$) in the scan plane to the coefficients of the first harmonic anisotropy through the spectral slope and ion speed ($v$):

$$A_1 = 2(\gamma+1)(V_R/v), \qquad B_1 = 2(\gamma+1)(V_N/v) \qquad (2)$$

We have calibrated our fitting procedure by comparison with the $V_R$ and $V_T$ components of plasma flow measured by Voyager 2 in the heliosheath using the Plasma Science instrument[5] (Fig. 2), which directly measures the solar wind velocity. The Plasma Science instrument on Voyager 1 failed in 1980. The comparison in Fig. 2 shows that, with the exception of periods of weak field-aligned ion streaming that we can readily identify in both the Voyager 2 and Voyager 1 data, the velocities derived from directional intensities of low-energy ions by the method described above is able to reproduce the solar wind velocity components quite well in the Low-Energy Charged Particle Instrument's scan plane. This justifies a posteriori our assumption that the particle distribution function is essentially isotropic in the plasma frame.

Orientations of the scan plane in its usual and rolled configuration are shown in Fig. 3a, b. The results of our Fourier analysis of the five Voyager 1 roll periods are presented in Fig. 3c–g. Figure 3h shows the intensities of 53–85-keV heliosheath protons arriving from the sunward and anti-sunward directions. The roll period dates, numbers of rolls per period and fit coefficients ($C_0$, $A_1$ and $B_1$) are given in columns 1–5 of Table 1. The errors in $C_0$, $A_1$ and $B_1$ are determined by propagating the Poisson statistical uncertainties in the sectored counting rates, which are shown as vertical error bars ($\pm 1\sigma$) about the mean in Fig. 3c–g, using the equations that express $C_0$, $A_1$ and $B_1$ as functions of the sectored rates. Alternative representations of $A_1$ and $B_1$ in terms of the first-order anisotropy amplitude $\xi_1$ and azimuth $\phi_1$ are given in columns 6 and 7 of Table 1, and the spectral power-law index $\gamma$ is in column 8. The plasma convection velocity components $V_R$ and $V_N$ (columns 9 and 10) implied by the first harmonic are given along with their uncertainties, which are calculated by propagating the errors in fit coefficients $C_0$, $A_1$ and $B_1$ using equation (2). The time-weighted averages of $V_R$ and $V_N$ over all five rolls are summarized in row 6 of those two columns. The five determinations of $V_N$ in the rolled configuration are plotted in Fig. 3j along those of $V_R$ and $V_{T'}$, the latter two components calculated using 26-d-averaged ion angular data taken in the usual (unrolled) configuration.

The averages of the velocity components and their uncertainties derived from the five roll periods are $\langle V_R \rangle = -14 \pm 14$ km s$^{-1}$ and $\langle V_N \rangle = +3 \pm 11$ km s$^{-1}$. The negative mean radial velocity during the rolls is consistent within errors with the more statistically significant result ($-6.5 \pm 4.1$ km s$^{-1}$) from the 26-d averages spanning the five roll periods in Fig. 3i, j (filled circles). The 26-d averages of the azimuthal flow ($V_{T'}$) have a larger and statistically significant negative value ($-25.8 \pm 3.8$ km s$^{-1}$). Although azimuthal flow is not the topic of this report, we do not wish its clear signature to pass unnoticed.

We offer several arguments that the time-weighted average for $V_N$ over five rolls is statistically consistent with zero, and moreover that it is small in an absolute sense. First, if our measurements of $V_N$ are consistent with zero, we would expect that roughly half of our roll period measurements would have $V_N > 0$ and half would have $V_N < 0$. Over the five spacecraft rolls, two had $V_N > 0$ and three had $V_N < 0$. Second, the Poisson error bars for each roll always bracket zero, giving no indication of a systematic non-zero flow. Third, as the Poisson distribution can be approximated by a Gaussian because the number of counts accumulated in each sector exceeds 250 (Table 1, column 3), our five-roll result $\langle V_N \rangle = +3 \pm 11$ km s$^{-1}$ implies only a 16% probability that $\langle V_N \rangle$ exceeds $+14$ km s$^{-1}$. For comparison, the distant upstream flow velocity of the local interstellar medium is $\sim$25 km s$^{-1}$. The solar radial vector to Voyager 1 is $\sim$30° offset from the upstream flow direction, that is, from the expected 'nose' of the heliosheath. At this angle, most steady-state models show a positive meridional flow that within the heliosheath is a significant fraction of the distant upstream value or even exceeds it (because of the constriction of plasma streamlines as they divert around the heliosheath). Our results give us 84% confidence that $\langle V_N \rangle$ is less than half of the distant upstream flow, even though our error bars are larger than our mean velocity ($3 \pm 8$ km s$^{-1}$). This is a drastic difference from the steady-state predictions.

We therefore conclude from our values (3 km s$^{-1} \ll 25$ km s$^{-1}$) that Voyager 1 is not at present close to the heliopause, at least in the form that it has been envisioned up to now. In fact, it has been in the transition region of weak radial ($V_R$) flow for over two years now (Fig. 3j), during which time it travelled an additional 7.5 AU outwards from the Sun. We do not know how much farther outwards the transition region extends, and the longer it lasts in time, the less likely it is to be dominated by a temporal effect of the expansion and contraction of the heliopause during the 11-year solar activity cycle[3]. However, a non-stationary solar wind should be included in any realistic model. In any case, any theories that predict a meridional flow velocity significantly outside of the Voyager 1 statistical limits ($-8$ km s$^{-1} < \langle V_N \rangle < 14$ km s$^{-1}$) should be reassessed, perhaps necessitating a new theoretical formulation of the interaction of the solar wind with the local interstellar medium.

**Received 6 March; accepted 25 July 2012.**

1. Krimigis, S. M., Roelof, E. C., Decker, R. B. & Hill, M. E. Zero outward flow velocity for plasma in a heliosheath transition layer. *Nature* **474,** 359–361 (2011).
2. Pogorelov, N. V., Borovikov, S. N., Zank, G. P. & Ogino, T. Three-dimensional features of the outer heliosphere due to coupling between the interstellar and interplanetary magnetic fields. III. The effects of solar rotation and activity cycle. *Astrophys. J.* **696,** 1478–1490 (2009).

3. Borovikov, S. N., Pogorelov, N. V., Burlaga, L. F. & Richardson, J. D. Plasma near the heliosheath: observations and modeling. *Astrophys. J.* **728,** L21–L26 (2011).
4. Gleeson, L. J. & Axford, W. I. The Compton-Getting effect. *Astrophys. Space Sci.* **2,** 431–437 (1968).
5. Richardson, J. D. & Wang, C. Plasma in the heliosheath: 3.5 years of observations. *Astrophys. J.* **734,** L21–L24 (2011).

**Acknowledgements** This work was supported at The Johns Hopkins University Applied Physics Laboratory by NASA contract NNN06AA01C. We thank J. Aiello for his assistance with our graphical presentation. We are grateful to the staff of the Voyager project for performing the Voyager 1 roll manoeuvres that made our analyses possible.

**Author Contributions** R.B.D. performed the data analysis and contributed to the text; S.M.K. contributed to the text; E.C.R. contributed to the text and provided theory interpretation; and M.E.H. analysed elemental composition.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.B.D. (robert.decker@jhuapl.edu).

LETTER

doi:10.1038/nature11434

# Flexible metal-oxide devices made by room-temperature photochemical activation of sol–gel films

Yong-Hoon Kim[1], Jae-Sang Heo[2], Tae-Hyeong Kim[1], Sungjun Park[3], Myung-Han Yoon[3,4], Jiwan Kim[1], Min Suk Oh[1], Gi-Ra Yi[5], Yong-Young Noh[6] & Sung Kyu Park[2]

**Amorphous metal-oxide semiconductors have emerged as potential replacements for organic and silicon materials in thin-film electronics. The high carrier mobility in the amorphous state, and excellent large-area uniformity, have extended their applications to active-matrix electronics, including displays, sensor arrays and X-ray detectors[1–7]. Moreover, their solution processability and optical transparency have opened new horizons for low-cost printable and transparent electronics on plastic substrates[8–13]. But metal-oxide formation by the sol–gel route requires an annealing step at relatively high temperature[2,14–19], which has prevented the incorporation of these materials with the polymer substrates used in high-performance flexible electronics. Here we report a general method for forming high-performance and operationally stable metal-oxide semiconductors at room temperature, by deep-ultraviolet photochemical activation of sol–gel films. Deep-ultraviolet irradiation induces efficient condensation and densification of oxide semiconducting films by photochemical activation at low temperature. This photochemical activation is applicable to numerous metal-oxide semiconductors, and the performance (in terms of transistor mobility and operational stability) of thin-film transistors fabricated by this route compares favourably with that of thin-film transistors based on thermally annealed materials. The field-effect mobilities of the photo-activated metal-oxide semiconductors are as high as 14 and 7 $\mathrm{cm^2\,V^{-1}\,s^{-1}}$ (with an $Al_2O_3$ gate insulator) on glass and polymer substrates, respectively; and seven-stage ring oscillators fabricated on polymer substrates operate with an oscillation frequency of more than 340 kHz, corresponding to a propagation delay of less than 210 nanoseconds per stage.**

During recent decades, solution-processed organic and inorganic semiconductors have been intensively investigated for realizing large-area flexible and printed electronics by continuous-solution processes[11,12]. Nevertheless, organic semiconductors still suffer from operational instability, and have relatively low carrier mobility for high-end applications. Some inorganic materials are too reactive to control in ambient conditions, and thus have had limited scope for large-scale fabrication. Recently, amorphous or polycrystalline metal-oxide semiconductors have been proposed as alternative channel materials, because they exhibit excellent optical transparency and good thin-film transistor (TFT) performance in ambient conditions[2,14–19]. Wet chemical, 'sol–gel' methods can be used to form high-quality oxide films, but such methods typically require a high-temperature annealing step, which is not compatible with conventional polymer substrates. Thus, for the full realization of flexible, large-scale, solution-processed metal-oxide electronics, it is important to understand the chemistry involved in sol–gel metal-oxide formation, and to apply this knowledge to the low-temperature synthesis of metal-oxide semiconducting films that are compatible with flexible polymer substrates and open-chamber, continuous processes.

We have developed a new photo-annealing method for forming amorphous metal-oxide semiconductors, and have examined its viability for producing large-area uniform devices and integrated circuits on polymer substrates. We use photochemical activation induced by deep-ultraviolet (DUV) light from a low-pressure mercury lamp in an inert atmosphere (to prevent reactive ozone formation) to achieve high degrees of sol–gel condensation and film densification in amorphous metal-oxide semiconductor systems including indium gallium zinc oxide (IGZO), indium zinc oxide (IZO) and indium oxide ($In_2O_3$). Our results suggest that DUV-assisted metal-oxide formation is a general route to prepare high-performance, solution-processed metal-oxide semiconductor films with only small amounts of extra heat supplied, permitting the use of thermally sensitive substrate materials.

To explain the formation of high-quality sol–gel semiconductor films by DUV irradiation, we propose the following mechanism, based on experimental data from ultraviolet–visible absorption spectroscopy, X-ray photoelectron spectroscopy, high-resolution transmission electron microscopy (HRTEM), Rutherford backscattering spectrometry and ellipsometry (Fig. 1 and Supplementary Fig. 1). When metal precursors for IGZO films are dissolved in 2-methoxyethanol (2-ME), and the resultant precursor solution is stirred at 75 °C for more than 12 h, a ligand exchange reaction occurs from nitrate/acetate to 2-methoxyethoxide or hydroxide, and condensation of metal alkoxides/hydroxides proceeds to form a partial network of metal–oxygen–metal (M–O–M) bonds in the solution. The as-spun films (25–35 nm thick) before DUV irradiation still contain a significant amount of residual organic components, as confirmed by a high carbon content in the film (Fig. 1b). Subsequently, when the as-spun film is exposed to DUV irradiation from the mercury lamp (main peaks at 184.9 nm (10%) and 253.7 nm (90%)) under nitrogen purging, high-energy DUV photons induce photochemical cleavage of alkoxy groups, and activate metal and oxygen atoms to facilitate M–O–M network formation (Fig. 1a, step 1, condensation). The efficiency of these DUV-assisted initial cleavage and condensation reactions is indicated by the rapid decrease of oxygen and carbon contents in the first 30 min of irradiation. Further irradiation induces a gradual removal of oxygen and carbon (and, thereby, near-complete condensation) and a transition to film densification (step 2, densification).

The degree of film densification is confirmed by comparing the areal densities and thicknesses of photo-annealed (P) and high-temperature, thermally annealed (T) IGZO films: $52.88 \times 10^{15}$ (P) versus $52.43 \times 10^{15}$ (T) atoms $\mathrm{cm^{-2}}$ from Rutherford backscattering spectrometry; and 7.1–9.70 (P) versus 7.1–10.26 (T) nm from HRTEM (lower limit) and ellipsometry (upper limit) measurements (Supplementary Fig. 1). Also, the atomic binding states, such as M–O bonding, in the photo-annealed film are similar to those in the thermally annealed film (Fig. 1c and Supplementary Fig. 2). We speculate that

---

[1] Flexible Display Research Center, Korea Electronics Technology Institute, Seongnam 463-816, Korea. [2] School of Electrical and Electronics Engineering, Chung-Ang University, Seoul 156-756, Korea. [3] School of Materials Science and Engineering, Gwangju Institute of Science and Technology, Gwangju 500-712, Korea. [4] Department of Nanobio Materials and Electronics, World Class University, Gwangju Institute of Science and Technology, Gwangju 500-712, Korea. [5] Department of Polymer Science and Engineering, Sungkyunkwan University, Suwon 440-746, Korea. [6] Department of Chemical Engineering, Hanbat National University, Daejeon 305-719, Korea.

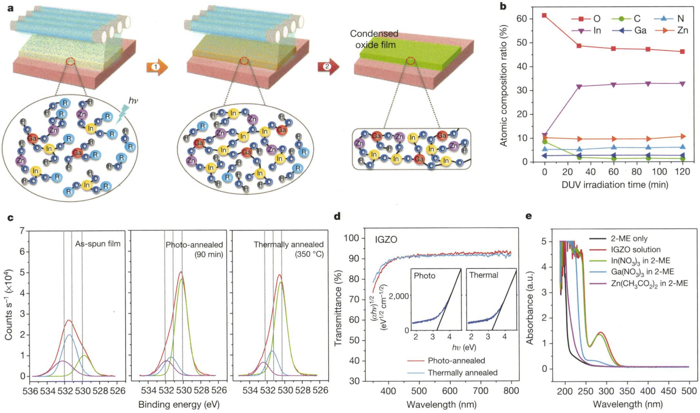


**Figure 1 | Photo-activation of solution-processed metal-oxide semiconductors by DUV. a**, Schemes showing condensation mechanism of metal-oxide precursors by DUV irradiation (*hν*). Light-blue shading denotes illumination from the low-pressure mercury lamp (blue cylinders). **b**, Atomic composition ratios of IGZO thin films as a function of DUV irradiation time. **c**, Red curves are X-ray photoelectron spectra (O(1*s*) peak) of as-spun, photo-annealed and thermally annealed IGZO films. (Deconvolution of the spectra shows the contributions of peaks at ~530.0 (green), ~531.0 (blue) and ~532.0 eV (purple) from, respectively: oxygen atoms in M–O–M lattice; oxygen atoms near oxygen vacancies and in M–OC bonds; and oxygen atoms in M–OH compounds.) **d**, Optical transmittance (main plot) and bandgap (insets) of thermally annealed and photo-annealed IGZO films on glass substrates. As shown by the tangent lines in the insets, the bandgaps of the photo-annealed and thermally annealed films are 3.23 and 3.22 eV, respectively. α, absorption coefficient. **e**, Light absorption characteristics of 2-ME, IGZO solution, and metal precursor solutions. a.u., arbitrary units.

such a high-degree of densification after 60 min is enabled by decomposition of organic residues (solvent molecules and residual alkoxy groups) by DUV-assisted photolysis and reorganization of M–O–M networks. The latter process is promoted by photochemical cleavage and rearrangement of disordered M–O–M networks without high-temperature annealing[20–24].

We have discovered that the DUV irradiation in our setup is accompanied by unintentional heating of the films up to ~150 °C (from the radiant heat of the lamp), and this temperature is maintained even after prolonged DUV irradiation (>120 min, >180–201 J cm$^{-2}$; Supplementary Fig. 3a). For comparison, metal-oxide films annealed at 150 °C without DUV treatment, or cooled on a cooling stage (40–70 °C) with DUV irradiation, showed almost no or low electrical performance, respectively (Fig. 2b and Supplementary Fig. 3d). All of these observations imply that near-complete condensation and densification of films requires both DUV photo-activation and the unintentional moderate heating. We suppose that this moderate heating provides extra thermal energy for the removal of volatile organic residues (2-ME has a boiling point of 124 °C), and for M–O–M network reorganization via efficient condensation and subsequent densification. Additionally, the measured optical transmittance and band gap of photo-annealed oxide films are very close to those of oxide films annealed at 350 °C (Fig. 1d, inset), and there is no apparent indication of DUV-induced metallic reduction[23].

Figure 1e shows ultraviolet–visible absorption spectra of precursor solutions for IGZO film preparation. For comparison, the absorption spectra of neat solvent (2-ME) and individual metal (In, Ga, Zn) precursor solutions are also shown. Unlike 2-ME, which shows minimal absorption at wavelengths of 225–350 nm, the solutions of $In(NO_3)_3 \cdot xH_2O$, $Ga(NO_3)_3 \cdot xH_2O$, and $Zn(CH_3CO_2)_2 \cdot 2H_2O$ in 2-ME exhibit strong light absorption below 260, 250, and 230 nm, respectively. As the mercury lamp has two main emission peaks at 253.7 and 184.9 nm, the photochemical activation of indium, gallium and zinc precursor molecules can be facilitated by DUV irradiation from the lamp.

Following successful application of the DUV photo-annealing method to IGZO thin films, we investigated its applicability to sol–gel films of other binary, ternary and quaternary oxide systems such as $In_2O_3$, IZO, zinc tin oxide (ZTO), and indium zinc tin oxide (IZTO). From preliminary tests with a simplified device architecture (on an $SiO_2$/Si wafer without channel isolation), we have concluded that the photo-annealing method can be applied to solution-based oxide systems except those using $ZnCl_2$ solution. This exception can be ascribed to the negligible DUV absorption in $ZnCl_2$ solution, leading to inefficient photochemical activation, cleavages and energy transfer by the high-energy DUV photons (Supplementary Fig. 4). Note that, whereas the IGZO TFTs photo-annealed in an $N_2$ atmosphere have shown excellent device characteristics, the performance of the devices photo-annealed in air is rather poor and unstable, despite an increase in substrate temperature to 180 °C (possibly due to the absence of $N_2$ purging; Supplementary Fig. 5). In air, the photo-activation efficiency by 184.9-nm emission from the mercury lamp is significantly attenuated, mainly owing to absorption by molecular oxygen ($O_2$)[22,23]. This causes insufficient photochemical cleavage of metal alkoxides and poor densification of the resultant film, leading to inactive TFT operation.

For further investigation of the effectiveness of DUV photo-activation for general oxide semiconductor preparation, we fabricated

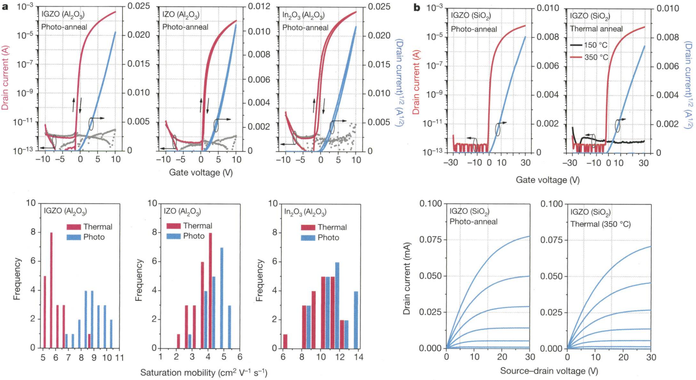


**Figure 2 | Transfer characteristics of photo-annealed IGZO, IZO and $In_2O_3$ TFTs using $Al_2O_3$ and $SiO_2$ gate dielectric, and comparison with thermally annealed devices.** **a**, Transfer characteristics and saturation mobility distribution of photo-annealed IGZO, IZO and $In_2O_3$ TFTs fabricated on glass with $Al_2O_3$ gate dielectric (~20 devices). The source–drain voltage, $V_{DS}$, is 10 V in all cases. In top panels, red curves are drain current; blue curves are (drain current)$^{1/2}$; and grey points are gate leakage current ($I_G$; values on left-hand axis). **b**, Transfer and output characteristics of photo-annealed and thermally annealed (150 and 350 °C) IGZO TFTs fabricated on $SiO_2$ (200 nm)/Si wafers. Red and blue curves as in **a**. The channel lengths and widths of all measured devices are 10 μm and 100 μm, respectively. In bottom panels, the seven curves are for source–gate voltages, $V_{GS}$, ranging from 0 to 30 V (bottom to top), in 5-V steps.

both thermally annealed and photo-annealed IGZO, IZO and $In_2O_3$ TFTs, and compared their performance. For the channel layer, the as-spun sol–gel films were photo-annealed in an $N_2$ atmosphere for 90–120 min, corresponding to an irradiation dose of 135–201 J cm$^{-2}$. Interestingly, despite the very small and gradual change in the atomic compositions after 60 min (Fig. 1b), the transistor mobilities of the formed films increase substantially after 90 min, with the best electrical properties and spatial uniformity achieved between 90 and 120 min of DUV photo-annealing (Supplementary Figs 5c and 6). These two distinct trends in atomic composition and electrical properties versus photo-activation time suggest that there are two separate stages of photo-activation: first, rapid chemical condensation, followed by gradual structural rearrangement and densification. The requirement for prolonged DUV exposure, with its accompanying moderate heating effect, may facilitate the second stage of photo-activation. Note that there is a distribution of optimal photo-activation times (Supplementary Figs 5c and 6), possibly due to uneven light intensity and/or power fluctuation of the mercury lamp currently installed in our photo-annealing apparatus.

Figure 2a shows the transfer characteristics of photo-annealed oxide TFTs with channel length and width of 10 and 100 μm, respectively, and with 35-nm-thick atomic-layer-deposited $Al_2O_3$ as a gate dielectric (138 nF cm$^{-2}$) on glass substrates. The photo-annealed TFTs have shown field-effect mobilities of $8.76 \pm 0.98$ cm$^2$ V$^{-1}$ s$^{-1}$ for IGZO, $4.43 \pm 0.59$ cm$^2$ V$^{-1}$ s$^{-1}$ for IZO, and $11.29 \pm 1.62$ cm$^2$ V$^{-1}$ s$^{-1}$ for $In_2O_3$. Compared with TFTs annealed at 350 °C, the photo-annealed

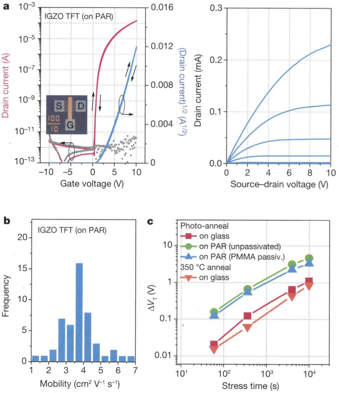


**Figure 3 | Electrical characteristics and bias stability of photo-annealed IGZO TFTs on flexible substrates.** **a**, Transfer and output characteristics of a photo-annealed IGZO TFT fabricated on a PAR substrate. Left panel: curves and grey points as in Fig. 2. Right panel: curves are for $V_{GS}$ ranging from 0 to 10 V (bottom to top), in 2-V steps. Channel length and width are 10 μm and 100 μm, respectively. **b**, Distribution of saturation mobilities of photo-annealed IGZO TFTs on PAR (49 devices). **c**, Threshold voltage shift, $\Delta V_T$, of IGZO TFTs under positive gate-bias stress ($V_{GS} = +5$ V, $V_{DS} = +0.1$ V). Glass substrates are unpassivated; PAR substrates are either unpassivated (green curve) or passivated with poly(methylmethacrylate) (blue curve).

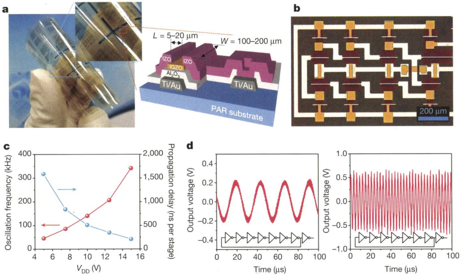


**Figure 4 | Characteristics of seven-stage ring oscillators fabricated on a PAR substrate by photo-annealing.** **a**, Optical micrographs and a schematic cross-section of photo-annealed IGZO TFTs and circuits on PAR. **b**, Optical micrograph of a seven-stage ring oscillator, with a $\beta$-ratio of 2 (see text for details of channel width/length ratios). Gate to source/drain overlap distance is 5 μm. **c**, Oscillation frequency (red) and per-stage propagation delay (blue) of seven-stage ring oscillator as a function of supply voltage, $V_{DD}$. **d**, Output waveforms of the seven-stage ring oscillator operating with supply voltages of 5 V (left panel) and 15 V (right panel), and oscillation frequencies of 45 and 341 kHz, respectively.

devices exhibit comparable or enhanced mobilities (Fig. 2a and Supplementary Fig. 7). Figure 2b shows the transfer and output characteristics of the room-temperature photo-annealed and high-temperature thermally annealed IGZO TFTs using thermally grown $SiO_2$ (200 nm) as a gate dielectric. The photo-annealed TFTs have shown field-effect mobilities as high as 2.64 $cm^2 V^{-1} s^{-1}$ (Supplementary Fig. 6b), which is also comparable to those of the thermally annealed devices at high temperature (350–500 °C)[15,18,25]. We speculate that the different semiconductor mobilities on $Al_2O_3$ and $SiO_2$ gate dielectrics may result from different values of the effective gate electric field (related to the gate insulator capacitance and applied gate bias), and the semiconductor–dielectric interface effect[2]. Nonetheless, these results show that photo-annealing is an alternative route to high-performance semiconductors based on solution-processed metal oxide films, even at room temperature.

To take full advantage of low-temperature photo-activation of metal-oxide semiconductors, we fabricated TFTs and circuits based on a solution-processed and photo-activated oxide semiconductor directly on commercially available polyarylate (PAR) film. The DUV irradiation induces a slight yellowing of the PAR substrate surface (optical transmission loss by 5–10%), but this DUV-induced colouration does not propagate beyond the topmost surface of PAR substrates, and the mechanical integrity of the film is minimally affected (Supplementary Fig. 9). Figure 3a, b shows typical device characteristics for TFTs made from photo-annealed IGZO on PAR substrates. The measured field-effect mobilities are centred at 3.77 $cm^2 V^{-1} s^{-1}$ (maximum value of ~7 $cm^2 V^{-1} s^{-1}$) with a narrow distribution (standard deviation of 1.02 $cm^2 V^{-1} s^{-1}$, from 49 devices). Also, the devices show excellent current on/off modulation, sub-threshold swing, and threshold voltage ($V_T$) values of $10^8$, 95.8 ± 20.8 mV per decade, and 2.70 ± 0.47 V, respectively.

We performed positive-gate-bias stress tests to verify the operational stability of photo-annealed IGZO TFTs in air, in dark conditions (Fig. 3c). Even without device packaging or passivation, the photo-annealed IGZO TFTs fabricated on glass substrates reveal outstanding operational stability, with a very small $V_T$ shift ($\Delta V_T$) of 1.12 V after a gate-bias stress time of 10,000 s (Supplementary Fig. 10). Note that the gate-bias stability is comparable to that of devices annealed at 350 °C under identical stress condition ($\Delta V_T$ of 0.86 V), and exceptionally low compared with that of previously reported devices based on solution-processed metal-oxide semiconductors[15,16,26–28]. In the case of photo-annealed IGZO TFTs on PAR substrates, the unpassivated and poly(methyl methacrylate)-passivated (~300-nm-thick) devices exhibit $\Delta V_T$ values of 4.5 and 3 V, respectively. More stable TFT characteristics on glass substrates may be attributed to the presence of fewer interfacial trap states at the interface between semiconductor and gate dielectric, possibly as a result of the low surface roughness of the dielectric layer[29] (Supplementary Fig. 11). To demonstrate device scalability, we fabricated seven-stage ring oscillator circuits on the PAR substrates (Fig. 4a, b). The room-temperature-fabricated IGZO TFTs on polymer substrates are typically enhancement-mode devices, and allow simple digital logic circuits without level shifting. The inverter in the ring oscillator had a $\beta$-ratio of 2 (channel width-to-length ratios $(W/L)_{\text{drive}} = 100\ \mu m/7\ \mu m$ and $(W/L)_{\text{load}} = 50\ \mu m/7\ \mu m$), with an overlap distance of 5 μm between the gate and source/drain electrodes. With a supply voltage of $V_{DD} = 15$ V, we measured an oscillation frequency greater than ~340 kHz, and corresponding propagation delay less than ~210 ns per stage (Fig. 4c, d).

We propose that DUV-assisted photochemistry approaches can open a new route for achieving high-performance, flexible and printed, metal-oxide thin-film electronic devices. Translation of this photo-annealing process to industrial applications may be helped by modifying the sol–gel solutions to include DUV-decomposable additives (fuels) and solvents, as well as by increasing the DUV energy density to boost the DUV-assisted photo-activation.

## METHODS SUMMARY

We prepared solutions for IGZO, IZO and $In_2O_3$ by dissolving indium nitrate hydrate, gallium nitrate hydrate and zinc acetate dihydrate in 2-ME (Supplementary Fig. 12). DUV photo-annealing was conducted by placing the as-spun films under a high-density DUV treatment system (UV253H, Filgen) under $N_2$ purging (film spacing 1–5 cm, 25–28 mW $cm^{-2}$). The light source is a low-pressure mercury lamp with two main emission peaks at 253.7 nm (90%) and 184.9 nm (10%).

For the fabrication of photo-annealed metal-oxide TFTs on glass substrates, we used 0.7-mm-thick glass (Eagle 2000, Samsung Corning Precision Glass). Gate electrodes were defined by patterning Ti/Au (3 nm/80 nm) or Mo (100 nm) layers. Gate dielectrics were 35-nm-thick $Al_2O_3$, deposited by atomic layer deposition at 100 °C over the gate-patterned substrates. For the channel layer, oxide solutions were spin-coated and photo-annealed in $N_2$ atmosphere. After subsequent patterning of the channel layer by wet etching, via holes and 100-nm-thick IZO source/drain electrodes were fabricated. For reference devices, spin-coated IGZO, IZO and $In_2O_3$ films were thermally annealed at 350 °C for 60 min on a hot plate in air.

For the fabrication of photo-annealed metal-oxide TFTs on polymer substrates, we used 200-μm-thick PAR films (A200HC, Ferrania Technologies), which have good dimensional stability. Other TFT fabrication processes are identical to those on glass substrates.

**Full Methods** and any associated references are available in the online version of the paper.

**Received 13 February; accepted 23 July 2012.**

1. Nomura, K. *et al.* Room-temperature fabrication of transparent flexible thin-film transistors using amorphous oxide semiconductors. *Nature* **432,** 488–492 (2004).
2. Kim, M. G., Kanatzidis, M. G., Facchetti, A. & Marks, T. J. Low-temperature fabrication of high-performance metal oxide thin-film electronics via combustion processing. *Nature Mater.* **10,** 382–388 (2011).
3. Facchetti, A. & Marks, T. J. *Transparent Electronics* (Wiley, 2010).
4. Dusastre, V. *Materials for Sustainable Energy* (World Scientific, 2010).
5. Kamiya, T., Nomura, K. & Hosono, H. Present status of amorphous In-Ga-Zn-O thin-film transistors. *Sci. Technol. Adv. Mater.* **11,** 044305 (2010).
6. Jeon, S. *et al.* Nanometer-scale oxide thin film transistor with potential for high-density image sensor applications. *ACS Appl. Mater. Interfaces* **3,** 1–6 (2011).
7. Kim, K. M. *et al.* Competitive device performance of low-temperature and all-solution-processed metal-oxide thin-film transistors. *Appl. Phys. Lett.* **99,** 242109 (2011).
8. Yang, S. *et al.* Low-temperature processed flexible In-Ga-Zn-O thin-film transistors exhibiting high electrical performance. *Electron. Dev. Lett.* **32,** 1692–1694 (2011).
9. Wager, J. F. Transparent electronics. *Science* **300,** 1245–1246 (2003).
10. Cao, Q. *et al.* Medium-scale carbon nanotube thin-film integrated circuits on flexible plastic substrates. *Nature* **454,** 495–500 (2008).
11. Klauk, H. *Organic Electronics: Materials, Manufacturing and Applications* (Wiley-VCH, 2006).
12. Briseno, A. L. *et al.* Patterning organic single-crystal transistor arrays. *Nature* **444,** 913–917 (2006).
13. Sekitani, T., Zschieschang, U., Klauk, H. & Someya, T. Flexible organic transistors and circuits with extreme bending stability. *Nature Mater.* **9,** 1015–1022 (2010).
14. Han, S. Y., Herman, G. S. & Chang, C. Low-temperature, high-performance, solution-processed indium oxide thin-film transistors. *J. Am. Chem. Soc.* **133,** 5166–5169 (2011).
15. Jeong, S., Ha, Y. G., Moon, J., Facchetti, A. & Marks, T. J. Role of gallium doping in dramatically lowering amorphous-oxide processing temperatures for solution-derived indium zinc oxide thin-film transistors. *Adv. Mater.* **22,** 1346–1350 (2010).
16. Meyers, S. T. *et al.* Aqueous inorganic inks for low-temperature fabrication of ZnO TFTs. *J. Am. Chem. Soc.* **130,** 17603–17609 (2008).
17. Kim, Y. H., Han, J. I. & Park, S. K. Effect of Zn/Tin composition ratio on the operational stability of solution-processed zinc tin oxide thin film transistors. *IEEE Electron Device Lett.* **33,** 50–52 (2012).
18. Kim, Y. H., Han, M. K., Han, J. I. & Park, S. K. Effect of metallic composition on electrical properties of solution-processed indium-gallium-zinc-oxide thin film transistors. *IEEE Trans. Electron. Dev.* **57,** 1009–1014 (2010).
19. Adamopoulos, G. *et al.* High-mobility low-voltage ZnO and Li-doped ZnO transistors based on $ZrO_2$ high-k dielectric grown by spray pyrolysis in ambient air. *Adv. Mater.* **23,** 1894–1898 (2011).
20. Van de Leest, R. E. UV photo-annealing of thin sol-gel films. *Appl. Surf. Sci.* **86,** 278–285 (1995).
21. Imai, H. in *Handbook of Sol-Gel Science and Technology: Processing, Characterization and Applications* Vol. 1 (ed. Sakka, S.) 639–650 (Kluwer, 2005).
22. Clark, T. Jr. *et al.* A new application of UV-ozone treatment in the preparation of substrate-supported, mesoporous thin films. *Chem. Mater.* **12,** 3879–3884 (2000).
23. Imai, H., Tominaga, A., Hirashima, H., Toki, M. & Asakuma, N. Ultraviolet-reduced reduction and crystallization of indium oxide films. *J. Appl. Phys.* **85,** 203–207 (1999).
24. Park, Y. M., Daniel, J., Heeney, M. & Salleo, A. Room-temperature fabrication of ultrathin oxide gate dielectrics for low-voltage operation of organic field-effect transistors. *Adv. Mater.* **23,** 971–974 (2011).
25. Hwang, S., Lee, J. H., Woo, C. H., Lee, J. Y. & Cho, H. K. Effect of annealing temperature on the electrical performances of solution-processed InGaZnO thin film transistors. *Thin Solid Films* **519,** 5146–5149 (2011).
26. Lim, W. *et al.* Improvement in bias stability of amorphous-$InGaZnO_4$ thin film transistors with $SiO_x$ passivation layers. *J. Vac. Sci. Technol. B* **28,** 116–119 (2010).
27. Son, K.-S. *et al.* Highly stable double-gate Ga-In-Zn-O thin-film transistor. *Electron Dev. Lett.* **31,** 812–814 (2010).
28. Cho, E. N., Kang, J. H., Kim, C. E., Moon, P. & Yun, I. Analysis of bias stress instability in amorphous InGaZnO thin-film transistors. *IEEE Trans. Device Mater. Reliab.* **11,** 112–117 (2011).
29. Choi, H. S. *et al.* Influence of Hf contents on interface state properties in *a*-HfInZnO thin-film transistors with $SiN_x/SiO_x$ gate dielectrics. *Appl. Phys. Lett.* **99,** 183502 (2011).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We acknowledge discussions with C.-I. Kim, S.-H. Song, H.-I. Kwon, B.-S. Bae, Y. Hong, S. Lim, J.-I. Han, M. J. Lee, A. Fenoglio and K.-H. Kim. This work was partially supported by Basic Science Research Program (no. 2010-0002623) and World-Class University Program (no. R31-10026) through a National Research Foundation of Korea (NRF) grant funded by the Ministry of Education, Science, and Technology.

**Author Contributions** S.K.P. designed the project and experiments; Y.-H.K., J.-S.H., T.-H.K., S.P., J.K., M.S.O., M.-H.Y. and S.K.P. carried out the experiments; S.K.P. and Y.-H.K. discussed and interpreted all the results; M.-H.Y., G.-R.Y. and Y.-Y.N. gave conceptual advice on the chemistry-related experiments and discussions. S.K.P., Y.-H.K., M.-H.Y. and G.-R.Y. wrote the manuscript, with S.K.P. the lead writer. All authors read and commented on the manuscript.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.K.P. (skpark@cau.ac.kr).

LETTER

doi:10.1038/nature11409

# Highly stretchable and tough hydrogels

Jeong-Yun Sun[1,2], Xuanhe Zhao[3], Widusha R. K. Illeperuma[1], Ovijit Chaudhuri[1], Kyu Hwan Oh[2], David J. Mooney[1,4], Joost J. Vlassak[1] & Zhigang Suo[1,5]

**Hydrogels are used as scaffolds for tissue engineering[1], vehicles for drug delivery[2], actuators for optics and fluidics[3], and model extracellular matrices for biological studies[4]. The scope of hydrogel applications, however, is often severely limited by their mechanical behaviour[5]. Most hydrogels do not exhibit high stretchability; for example, an alginate hydrogel ruptures when stretched to about 1.2 times its original length. Some synthetic elastic hydrogels[6,7] have achieved stretches in the range 10–20, but these values are markedly reduced in samples containing notches. Most hydrogels are brittle, with fracture energies of about $10\,\mathrm{J\,m^{-2}}$ (ref. 8), as compared with $\sim 1{,}000\,\mathrm{J\,m^{-2}}$ for cartilage[9] and $\sim 10{,}000\,\mathrm{J\,m^{-2}}$ for natural rubbers[10]. Intense efforts are devoted to synthesizing hydrogels with improved mechanical properties[11–18]; certain synthetic gels have reached fracture energies of $100$–$1{,}000\,\mathrm{J\,m^{-2}}$ (refs 11, 14, 17). Here we report the synthesis of hydrogels from polymers forming ionically and covalently crosslinked networks. Although such gels contain ~90% water, they can be stretched beyond 20 times their initial length, and have fracture energies of $\sim 9{,}000\,\mathrm{J\,m^{-2}}$. Even for samples containing notches, a stretch of 17 is demonstrated. We attribute the gels' toughness to the synergy of two mechanisms: crack bridging by the network of covalent crosslinks, and hysteresis by unzipping the network of ionic crosslinks. Furthermore, the network of covalent crosslinks preserves the memory of the initial state, so that much of the large deformation is removed on unloading. The unzipped ionic crosslinks cause internal damage, which heals by re-zipping. These gels may serve as model systems to explore mechanisms of deformation and energy dissipation, and expand the scope of hydrogel applications.**

Certain synthetic hydrogels exhibit exceptional mechanical behaviour. A hydrogel containing slide-ring polymers can be stretched to more than 10 times its initial length[6]; a tetra-poly(ethylene glycol) gel has a strength of ~2.6 MPa (ref. 7). These gels deform elastically. An elastic gel is known to be brittle and notch-sensitive; that is, the stretchability and strength decrease markedly when samples contain notches, or any other features that cause inhomogeneous deformation[19]. A gel can be made tough and notch-insensitive by introducing energy-dissipating mechanisms. For example, a fracture energy of $\sim 1{,}000\,\mathrm{J\,m^{-2}}$ is achieved with a double-network gel, in which two networks—one with short chains, and the other with long chains—are separately crosslinked by covalent bonds[11]. When the gel is stretched, the short-chain network ruptures and dissipates energy[20]. But the rupture of the short-chain network causes permanent damage. After the first loading, the gel does not recover from this damage; thus, on subsequent loadings, the fracture energy is much reduced[21]. To enable recoverable energy-dissipating mechanisms, several recent works have replaced the sacrificial covalent bonds with non-covalent bonds. In a gel with a copolymer of triblock chains, for example, the end blocks of different chains form glassy domains, and the midblocks of different chains form ionic crosslinks[22]. When the gel is stretched, the glassy domains remain intact, while the ionic crosslinks break and dissipate energy. The ionic crosslinks then re-form during a time interval after the first loading[22]. Recoverable energy dissipation can also be effected by hydrophobic associations[17,18]. When a gel made with hydrophobic bilayers in a hydrophilic polymer network is stretched, the bilayers dissociate and dissipate energy; on unloading, the bilayers re-assemble, leading to recovery[17]. However, previous studies along these lines have demonstrated fracture energy comparable to, or lower than, that of the double-network gels.

We have synthesized extremely stretchable and tough hydrogels by mixing two types of crosslinked polymer: ionically crosslinked alginate, and covalently crosslinked polyacrylamide (Fig. 1). An alginate chain comprises mannuronic acid (M unit) and guluronic acid (G unit), arranged in blocks rich in G units, blocks rich in M units, and blocks of alternating G and M units. In an aqueous solution, the G blocks in different alginate chains form ionic crosslinks through divalent cations (for example, $\mathrm{Ca^{2+}}$), resulting in a network in water—an alginate hydrogel. By contrast, in a polyacrylamide hydrogel, the polyacrylamide chains form a network by covalent crosslinks. We dissolved powders of alginate and acrylamide in deionized water. (Unless otherwise stated, the water content was fixed at 86 wt %.) We added ammonium persulphate as a photo-initiator for polyacrylamide, and N,N-methylenebisacrylamide as the crosslinker for polyacrylamide. After degassing the solution in a vacuum chamber, we added N,N,N′,N′-tetramethylethylenediamine, at 0.0025 the weight of acrylamide, as the crosslinking accelerator for polyacrylamide, and calcium sulphate slurry ($\mathrm{CaSO_4{\cdot}2H_2O}$) as the ionic crosslinker for alginate. We poured the solution into a glass mould measuring $75.0 \times 150.0 \times 3.0\,\mathrm{mm^3}$, covered with a 3-mm-thick glass plate. The gel was cured in one step with ultraviolet light for 1 hour (with 8 W power and 254 nm wavelength at 50 °C), and was then left in a humid box for 1 day to stabilize the reactions. After the curing step, we took the gel out of the humid box, and removed water on its surfaces using $\mathrm{N_2}$ gas for 1 minute.

The gel was glued to two polystyrene clamps, resulting in specimens measuring $75.0 \times 5.0 \times 3.0\,\mathrm{mm^3}$. All mechanical tests were performed in air, at room temperature, using a tensile machine with a 500-N load cell. In both loading and unloading, the rate of stretch was kept constant at $2\,\mathrm{min^{-1}}$. We stretched an alginate–polyacrylamide hybrid gel to >20 times its original length without rupture (Fig. 2a,b). The hybrid gel was also extremely notch-insensitive. When we cut a notch into the gel (Fig. 2c) and then pulled it to a stretch of 17, the notch was dramatically blunted and remained stable (Fig. 2d). At a critical applied stretch, a crack initiated at the front of the notch, and ran rapidly through the entire sample (Supplementary Movie 1). Large, recoverable deformation is demonstrated by dropping a metal ball on a membrane of the gel fixed by circular clamps (Supplementary Movie 2). On hitting the membrane, the ball stretched the membrane greatly and then bounced back. The membrane remained intact, vibrated, and recovered its initial flat configuration after the vibration was damped out. A ball with greater kinetic energy, however, caused the membrane to rupture after large deformation (Supplementary Movie 3).

The extremely stretchable hybrid gels are even more remarkable when compared with their parents, the alginate and polyacrylamide

---

[1] School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA. [2] Department of Material Science and Engineering, Seoul National University, Seoul 151-742, South Korea. [3] Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708, USA. [4] Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, Massachusetts 02138, USA. [5] Kavli Institute for Bionano Science and Technology, Harvard University, Cambridge, Massachusetts 02138, USA.

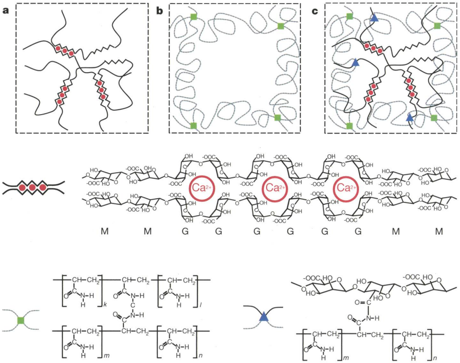


**Figure 1 | Schematics of three types of hydrogel. a**, In an alginate gel, the G blocks on different polymer chains form ionic crosslinks through $Ca^{2+}$ (red circles). **b**, In a polyacrylamide gel, the polymer chains form covalent crosslinks through N,N-methylenebisacrylamide (MBAA; green squares). **c**, In an alginate–polyacrylamide hybrid gel, the two types of polymer network are intertwined, and joined by covalent crosslinks (blue triangles) between amine groups on polyacrylamide chains and carboxyl groups on alginate chains. Materials used were as follows: alginate (FMC Biopolymer, LF 20/40); acrylamide (Sigma, A8887); ammonium persulphate (Sigma, A9164); MBAA (Sigma, M7279); N,N,N′,N′-tetramethylethylenediamine (Sigma, T7024); $CaSO_4 \cdot 2H_2O$ (Sigma, 31221); ultraviolet lamp (Hoefer, UVC 500).

gels (Fig. 3a). The amounts of alginate and acrylamide in the hybrid gels were kept the same as those in the alginate gel and polyacrylamide gel, respectively. When the stretch was small, the elastic modulus of the hybrid gel was 29 kPa, which is close to the sum of the elastic moduli of the alginate and polyacrylamide gels (17 kPa and 8 kPa, respectively). The stress and stretch at rupture were, respectively, 156 kPa and 23 for the hybrid gel, 3.7 kPa and 1.2 for the alginate gel, and 11 kPa and 6.6 for the polyacrylamide gel. Thus, the properties at rupture of the hybrid gel far exceeded those of either of its parents.

Hybrid gels dissipate energy effectively, as shown by pronounced hysteresis. The area between the loading and unloading curves of a gel gives the energy dissipated per unit volume (Fig. 3b). The alginate gel exhibited pronounced hysteresis and retained significant permanent deformation after unloading. In contrast, the polyacrylamide gel showed negligible hysteresis, and the sample fully recovered its original length after unloading. The hybrid gel also showed pronounced hysteresis, but the permanent deformation after unloading was significantly smaller than that of the alginate gel. The pronounced hysteresis and relatively small permanent deformation of the hybrid gel were further demonstrated by loading several samples to large values of stretch before unloading (Fig. 3c).

After the first loading and unloading, the hybrid gel was much weaker if the second loading was applied immediately, and recovered somewhat if the second loading was applied 1 day later (Fig. 3d and Supplementary Fig. 1). We loaded a sample of the hybrid gel to a stretch of 7, and then unloaded the gel to zero force. The sample was

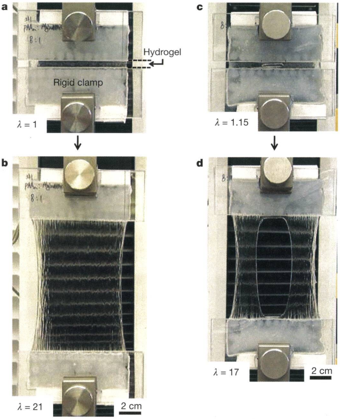


**Figure 2 | The hybrid gel is highly stretchable and notch-insensitive. a**, A strip of the undeformed gel was glued to two rigid clamps. **b**, The gel was stretched to 21 times its initial length in a tensile machine (Instron model 3342). The stretch, $\lambda$, is defined by the distance between the two clamps when the gel is deformed, divided by the distance when the gel is undeformed. **c**, A notch was cut into the gel, using a razor blade; a small stretch of 1.15 was used to make the notch clearly visible. **d**, The gel containing the notch was stretched to 17 times its initial length. The alginate/acrylamide ratio was 1:8. The weight of the covalent crosslinker, MBAA, was fixed at 0.0006 that of acrylamide; the weight of the ionic crosslinker, $CaSO_4$, was fixed at 0.1328 that of alginate.

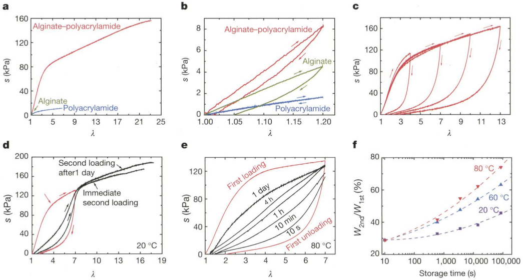


**Figure 3 | Mechanical tests under various conditions.** **a**, Stress–stretch curves of the three types of gel, each stretched to rupture. The nominal stress, *s*, is defined as the force applied on the deformed gel, divided by the cross-sectional area of the undeformed gel. **b**, The gels were each loaded to a stretch of 1.2, just below the value that would rupture the alginate gel, and were then unloaded. **c**, Samples of the hybrid gel were subjected to a cycle of loading and unloading of varying maximum stretch. **d**, After the first cycle of loading and unloading (red curve), one sample was reloaded immediately, and the other sample was reloaded after 1 day (black curves, as labelled). **e**, Recovery of samples stored at 80 °C for different durations, as labelled. **f**, The work of the second loading, $W_{2nd}$, normalized by that of the first loading, $W_{1st}$, measured for samples stored for different durations at different temperatures. The alginate/acrylamide ratio was 1:8 for **a** and **b**, and 1:6 for **c**–**f**. Weights of crosslinkers were fixed as described in Fig. 2 legend.

then sealed in a polyethylene bag and submerged in mineral oil to prevent water from evaporating, and stored in a fixed-temperature bath for a prescribed time. The sample was then taken out of storage and its stress–stretch curve was measured again at room temperature. The internal damage was much better healed by storing the gel at an elevated temperature for some time before reloading (Fig. 3e and Supplementary Fig. 2). After storing at 80 °C for 1 day, the work on reloading was recovered to 74% of that of the first loading (Fig. 3f).

We prepared gels containing various proportions of alginate and acrylamide to study why the hybrids were much more stretchable and stronger than either of their parents. When the proportion of acrylamide was increased, the elastic modulus of the hybrid gel decreased (Fig. 4a). However, the critical stretch at rupture reached a maximum at 89 wt % acrylamide. A similar trend was observed for samples with notches (Fig. 4c). The fracture energy reached a maximum value of 8,700 J m$^{-2}$ at 86 wt % acrylamide (Fig. 4d). The densities of ionic and covalent crosslinks also strongly affect the mechanical behaviour of the hybrid gels (Supplementary Figs 3, 4), as well as that of pure alginate gels (Supplementary Fig. 5) and pure polyacrylamide gels (Supplementary Fig. 6).

Our experimental findings provide insight into the mechanisms of deformation and energy dissipation in these gels. When an unnotched hybrid gel is subjected to a small stretch, the elastic modulus of the hybrid gel is nearly the sum of those of the alginate and polyacrylamide gels. This behaviour is also suggested by viscoelastic moduli determined for the hybrid and pure gels (Supplementary Fig. 7). Thus, in the hybrid gel the alginate and polyacrylamide chains both bear loads. Moreover, alginate is finely and homogeneously dispersed in the hybrid gel, as demonstrated by using fluorescent alginate and by measuring local elastic modulus with atomic force microscopy (Supplementary Fig. 8). The load sharing of the two networks may be achieved by entanglements of the polymers, and by possible covalent crosslinks formed between the amine groups on polyacrylamide chains and the carboxyl groups on alginate chains (Fig. 1, Supplementary Figs 9, 10). As the stretch increases, the alginate network unzips progressively[23], while the polyacrylamide network remains intact, so that the hybrid gel exhibits pronounced hysteresis and little permanent deformation. As only the ionic crosslinks are broken, and the alginate chains themselves remain intact, the ionic crosslinks can re-form, leading to the healing of the internal damage.

The giant fracture energy of the hybrid gel is remarkable, considering that its parents—the alginate and polyacrylamide gels—have fracture energies of 10–250 J m$^{-2}$ (Supplementary Figs 5, 6). The relatively low fracture energy of a hydrogel comprising a single network with covalent crosslinks is understood in terms of the Lake–Thomas model[8]. When the gel contains a notch and is stretched, the deformation is inhomogeneous; the network directly ahead of the notch is stretched more than elsewhere (Supplementary Fig. 11). For the notch

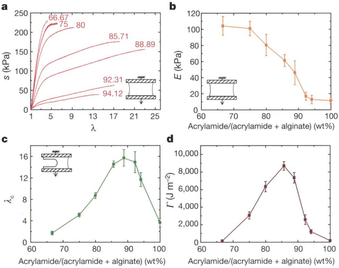


**Figure 4 | Composition greatly affects behaviour of the hybrid gel.** **a**, Stress–strain curves of gels of various weight ratios of acrylamide to (acrylamide plus alginate), as labelled. Each test was conducted by pulling an unnotched sample to rupture. **b**, Elastic moduli calculated from stress–strain curves, plotted against weight ratio. **c**, Critical stretch, $\lambda_c$, for notched gels of various weight ratios, measured by pulling the gels to rupture. **d**, Fracture energy, $\Gamma$, as a function of weight ratio. Weights of crosslinkers were fixed as described in Fig. 2 legend. Error bars show standard deviation; sample size $n = 4$.

to turn into a running crack, only the chains directly ahead of the notch need to break. Once a chain breaks, the energy stored in the entire chain is dissipated. In the ionically crosslinked alginate, fracture proceeds by unzipping ionic crosslinks and pulling out chains[24]. After one pair of G blocks unzip, the high stress shifts to the neighbouring pair of G blocks and causes them to unzip also (Supplementary Fig. 11). For the notch in the alginate gel to turn into a running crack, only the alginate chains crossing the crack plane need to unzip, leaving the network elsewhere intact. In both polyacrylamide gel and alginate gel, rupture results from localized damage, leading to small fracture energies.

That a tough material can be made of brittle constituents is reminiscent of transformation-toughening ceramics, and of composites made of ceramic fibres and ceramic matrices. The toughness of the hybrid gel can be understood by adapting a model well studied for toughened ceramics[25] and for gels of double networks of covalent crosslinks[26,27]. When a notched hybrid gel is stretched, the polyacrylamide network bridges the crack and stabilizes deformation, enabling the alginate network to unzip over a large region of the gel (Supplementary Fig. 11). The unzipping of the alginate network, in its turn, reduces the stress concentration of the polyacrylamide network ahead of the notch. The model highlights the synergy of the two toughening mechanisms: crack bridging and background hysteresis.

The idea that gels can be toughened by mixing weak and strong bonds has been exploited in several ways, including hydrophobic associations[18], particle-filled gels[7,15] and supramolecular chemistry[17,22]. The fracture energy of the alginate–polyacrylamide hybrid gel, however, is much larger than previously reported values[14,17,20,28] for tough synthetic gels (100–1,000 $\mathrm{J\,m^{-2}}$), a finding that we attribute to how the alginate network unzips. Each alginate chain contains a large number of G blocks, many of which form ionic crosslinks with G blocks on other chains when enough $Ca^{2+}$ ions are present[1]. When the hybrid gel is stretched, the polyacrylamide network remains intact and stabilizes the deformation, while the alginate network unzips progressively, with closely spaced ionic crosslinks unzipping at a small stretch, followed by more and more widely spaced ionic crosslinks unzipping as the stretch increases.

Because of the large magnitude of the fracture energy and the pronounced blunting of the notches, we ran a large number of experiments to determine the fracture energy, using three types of specimen, as well as changing the size of the specimens (Supplementary Figs 12–16). The experiments showed that the measured fracture energy is independent of the shape and size of the specimens.

Our data suggest that the fracture energy of hydrogels can be greatly increased by combining weak and strong crosslinks. The combination of relatively high stiffness, high toughness and recoverability of stiffness and toughness, along with an easy method of synthesis, make these materials ideal candidates for further investigation. Further development is needed to relate macroscopically observed mechanical behaviour to microscopic parameters. Many types of weak and strong molecular integrations can be used, making hybrid gels of various kinds a fertile area of research. In many applications, the use of hydrogels is often severely limited by their mechanical properties. For example, the poor mechanical stability of hydrogels used for cell encapsulation often leads to unintended cell release and death[29], and low toughness limits the durability of contact lenses[30]. Hydrogels of superior stiffness, toughness, stretchability and recoverability will improve the performance in these applications, and will probably open up new areas of application for this class of materials.

**Received 16 February; accepted 10 July 2012.**

1. Lee, K. Y. & Mooney, D. J. Hydrogels for tissue engineering. *Chem. Rev.* **101,** 1869–1879 (2001).
2. Qiu, Y. & Park, K. Environment-sensitive hydrogels for drug delivery. *Adv. Drug Deliv. Rev.* **53,** 321–339 (2001).
3. Dong, L., Agarwal, A. K., Beebe, D. J. & Jiang, H. R. Adaptive liquid microlenses activated by stimuli-responsive hydrogels. *Nature* **442,** 551–554 (2006).
4. Discher, D. E., Mooney, D. J. & Zandstra, P. W. Growth factors, matrices, and forces combine and control stem cells. *Science* **324,** 1673–1677 (2009).
5. Calvert, P. Hydrogels for soft machines. *Adv. Mater.* **21,** 743–756 (2009).
6. Okumura, Y. & Ito, K. The polyrotaxane gel: a topological gel by figure-of-eight cross-links. *Adv. Mater.* **13,** 485–487 (2001).
7. Haraguchi, K. & Takehisa, T. Nanocomposite hydrogels: a unique organic-inorganic network structure with extraordinary mechanical, optical and swelling/de-swelling properties. *Adv. Mater.* **14,** 1120–1124 (2002).
8. Lake, G. J. & Thomas, A. G. The strength of highly elastic materials. *Proc. R. Soc. A* **300,** 108–119 (1967).
9. Simha, N. K., Carlson, C. S. & Lewis, J. L. Evaluation of fracture toughness of cartilage by micropenetration. *J. Mater. Sci. Mater. Med.* **14,** 631–639 (2003).
10. Lake, G. J. Fatigue and fracture of elastomers. *Rubber Chem. Technol.* **68,** 435–460 (1995).
11. Gong, J. P., Katsuyama, Y., Kurokawa, T. & Osada, Y. Double-network hydrogels with extremely high mechanical strength. *Adv. Mater.* **15,** 1155–1158 (2003).
12. Huang, T. *et al.* A novel hydrogel with high mechanical strength: a macromolecular microsphere composite hydrogel. *Adv. Mater.* **19,** 1622–1626 (2007).
13. Sakai, T. *et al.* Design and fabrication of a high-strength hydrogel with ideally homogeneous network structure from tetrahedron-like macromonomers. *Macromolecules* **41,** 5379–5384 (2008).
14. Seitz, M. E. *et al.* Fracture and large strain behavior of self-assembled triblock copolymer gels. *Soft Matter* **5,** 447–456 (2009).
15. Lin, W.-C., Fan, W., Marcellan, A., Hourdet, D. & Creton, C. Large strain and fracture properties of poly(dimethylacrylamide)/silica hybrid hydrogels. *Macromolecules* **43,** 2554–2563 (2010).
16. Wang, Q. G. *et al.* High-water-content mouldable hydrogels by mixing clay and a dendritic molecular binder. *Nature* **463,** 339–343 (2010).
17. Haque, M. A., Kurokawa, T., Kamita, G. & Gong, J. P. Lamellar bilayers as reversible sacrificial bonds to toghen hydrogel: hysteresis, self-recovery, fatigue resistance, and crack blunting. *Macromolecules* **44,** 8916–8924 (2011).
18. Tuncaboylu, D. C., Sari, M., Oppermann, W. & Okay, O. Tough and self-healing hydrogels formed via hydrophobic interactions. *Macromolecules* **44,** 4997–5005 (2011).
19. Hui, C.-Y., Jagota, A., Bennison, S. J. & Londono, J. D. Crack blunting and the strength of soft elastic solids. *Proc. R. Soc. Lond. A* **459,** 1489–1516 (2003).
20. Yu, Q. M., Tanaka, Y., Furukawa, H., Kurokawa, T. & Gong, J. P. Direct observation of damage zone around crack tips in double-network gels. *Macromolecules* **42,** 3852–3855 (2009).
21. Webber, R. E., Creton, C., Brown, H. R. & Gong, J. P. Large strain hysteresis and mullins effect of tough double-network hydrogels. *Macromolecules* **40,** 2919–2927 (2007).
22. Henderson, K. J., Zhou, T. C., Otim, K. J. & Shull, K. R. Ionically cross-linked triblock copolymer hydrogels with high strength. *Macromolecules* **43,** 6193–6201 (2010).
23. Kong, H. J., Wong, E. & Mooney, D. J. Independent control of rigidity and toughness of polymeric hydrogels. *Macromolecules* **36,** 4582–4588 (2003).
24. Baumberger, T. & Ronsin, O. From thermally activated to viscosity controlled fracture of biopolymer hydrogels. *J. Chem. Phys.* **130,** 061102 (2009).
25. Evans, A. G. Perspective on the development of high-toughness ceramics. *J. Am. Ceram. Soc.* **73,** 187–206 (1990).
26. Brown, H. R. A model of fracture of double network gels. *Macromolecules* **40,** 3815–3818 (2007).
27. Tanaka, Y. A local damage model for anomalous high toughness of double-network gels. *Europhys. Lett.* **78,** 56005 (2007).
28. Jackson, A. P. Measurement of the fracture toughness of some contact lens hydrogels. *Biomater.* **11,** 403–407 (1990).
29. Hernández, R. M., Orive, G., Murua, A. & Pedraz, J. L. Microcapsules and microcarriers for in situ cell delivery. *Adv. Drug Deliv. Rev.* **62,** 711–730 (2010).
30. Maldonado-Codina, C. & Efron, N. Impact of manufacturing technology and material composition on the mechanical properties of hydrogel contact lenses. *Ophthalmic Physiol. Opt.* **24,** 551–561 (2004).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** The work at Harvard was supported by ARO (W911NF-09-1-0476), NSF (CMMI-0800161), DARPA (W911NF-10-1-0113), NIH (R37 DE013033) and MRSEC (DMR-0820484). X.Z. acknowledges the support of the NSF Research Triangle MRSEC (DMR-1121107) and Haythornthwaite Research Initiation grants. K.H.O. is supported by the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science and Technology (R11-2005-065). Z.S. acknowledges a sabbatical leave at the Karlsruhe Institute of Technology funded by the Alexander von Humboldt Award and by Harvard University.

**Author Contributions** J.-Y.S., X.Z., W.R.K.I., D.J.M., J.J.V. and Z.S. designed the study and interpreted the results. X.Z. developed the protocol for fabrication of the gels and prepared initial samples. J.-Y.S. and W.R.K.I. improved the protocol, and performed mechanical tests and recovery tests. J.-Y.S. obtained Fourier transform infrared spectra and performed thermogravimetric analysis. O.C. and J.-Y.S. conducted the experiment with fluorescent alginate and that using the atomic force microscope. K.H.O. contributed to the discussion of results. J.-Y.S., W.R.K.I. and Z.S. wrote the manuscript. All authors commented on the manuscript.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Z.S. (suo@seas.harvard.edu).

LETTER

doi:10.1038/nature11392

# Activation of old carbon by erosion of coastal and subsea permafrost in Arctic Siberia

J. E. Vonk[1]†*, L. Sánchez-García[1]†*, B. E. van Dongen[1]†, V. Alling[1]†, D. Kosmach[2], A. Charkin[2], I. P. Semiletov[2,3], O. V. Dudarev[2], N. Shakhova[2,3], P. Roos[4], T. I. Eglinton[5], A. Andersson[1] & Ö. Gustafsson[1]

**The future trajectory of greenhouse gas concentrations depends on interactions between climate and the biogeosphere[1,2]. Thawing of Arctic permafrost could release significant amounts of carbon into the atmosphere in this century[3]. Ancient Ice Complex deposits outcropping along the ~7,000-kilometre-long coastline of the East Siberian Arctic Shelf (ESAS)[4,5], and associated shallow subsea permafrost[6,7], are two large pools of permafrost carbon[8], yet their vulnerabilities towards thawing and decomposition are largely unknown[9–11]. Recent Arctic warming is stronger than has been predicted by several degrees, and is particularly pronounced over the coastal ESAS region[12,13]. There is thus a pressing need to improve our understanding of the links between permafrost carbon and climate in this relatively inaccessible region. Here we show that extensive release of carbon from these Ice Complex deposits dominates (57 ± 2 per cent) the sedimentary carbon budget of the ESAS, the world's largest continental shelf, overwhelming the marine and topsoil terrestrial components. Inverse modelling of the dual-carbon isotope composition of organic carbon accumulating in ESAS surface sediments, using Monte Carlo simulations to account for uncertainties, suggests that 44 ± 10 teragrams of old carbon is activated annually from Ice Complex permafrost, an order of magnitude more than has been suggested by previous studies[14]. We estimate that about two-thirds (66 ± 16 per cent) of this old carbon escapes to the atmosphere as carbon dioxide, with the remainder being re-buried in shelf sediments. Thermal collapse and erosion of these carbon-rich Pleistocene coastline and seafloor deposits may accelerate with Arctic amplification of climate warming[2,13].**

The large magnitude of shallow permafrost carbon pools relative to the atmospheric pools of carbon dioxide ($\sim$760 Pg) and methane ($\sim$3.5 Pg) suggests that carbon release from thawing permafrost has the potential to affect large-scale carbon cycling. Arctic permafrost can be divided into three main compartments: terrestrial (tundra and taiga) permafrost ($\sim$1,000 Pg C)[8], Ice Complex (coastal and inland) permafrost ($\sim$400 Pg C)[4,8] and subsea permafrost ($\sim$1,400 Pg C)[6,7]. Even without considering subsea permafrost, the carbon held in the top few metres of the pan-arctic permafrost constitutes approximately half of the global soil organic carbon pool[8].

Investigations of Arctic greenhouse gas releases have focused on terrestrial permafrost systems[4,9,15], and only recently on subsea permafrost[6,7,16,17], with a notable scarcity of studies on the thawing permafrost outcropping along the Arctic coast. In particular, the extensive coastline of the Eastern Siberian Sea (ESS) is dominated by exposed tall bluffs comprising ice-rich, fine-grained Ice Complex deposits (Fig. 1a). The origin of the $\sim$1-million-km$^2$ deposits (with average depth 25 m) dominating northeastern Siberia (and parts of Alaska and northwestern Canada) is under some debate, but this Pleistocene material is quite distinct from peat and mineral soil of other Arctic permafrost[4,5]. These relict soils of the steppe-tundra ecosystem have high carbon contents (1–5%)[4,5]. The export of organic carbon from the eroding ESAS Ice Complex is presently estimated at 4 Tg yr$^{-1}$ (ref. 14), yet it has also been proposed that erosion from the Lena Delta coastline alone might contribute this amount[18]. Clearly, large uncertainties remain regarding the magnitude of eroded carbon export from land to the shelf.

The extensive coastal exposure of the Ice Complex deposits (ICD) makes them potentially more vulnerable than other terrestrial permafrost; ICD retreat rates are 5–7 times higher than those of other coastal permafrost bodies[18]. A destructive thaw-erosion process brought on by thermal collapse of the coastline promotes surface

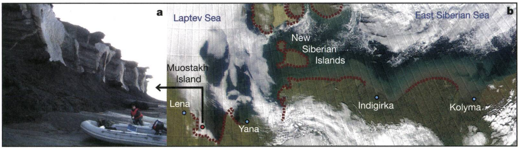


**Figure 1 | Erosion of Ice Complex deposits on the East Siberian Arctic Shelf.** **a**, Eroding, carbon-rich Ice Complex coast on Muostakh Island in the southeastern Laptev Sea. **b**, Erosion-induced turbidity clouds envelop several thousand kilometres of East Siberian Sea coastal waters. Note the rounded shorelines of northeastern Siberia, indicative of coastal erosion. Red dashed line shows areas of intensive ongoing erosion. (Satellite image of 24 August 2000, available at http://visibleearth.nasa.gov.)

---

[1]Department of Applied Environmental Science (ITM) and the Bert Bolin Centre for Climate Research, Stockholm University, Svante Arrhenius väg 8, SE-11418, Stockholm, Sweden. [2]Pacific Oceanological Institute, Russian Academy of Sciences, ul. Baltiiskaya 43, 690041, Vladivostok, Russia. [3]International Arctic Research Center, University of Alaska, PO Box 757335, Fairbanks, Alaska 99775-7335, USA. [4]Risø National Laboratory for Sustainable Energy, Frederiksborgvej 399, 4000, Roskilde, Denmark. [5]Geological Institute, ETH-Zürich, Sonneggstrasse 5, CH-8092, Zürich, Switzerland. †Present addresses: Geological Institute, ETH-Zürich, Sonneggstrasse 5, CH-8092, Zürich, Switzerland (J.E.V.); Catalan Institute of Climate Sciences (IC3), C/Doctor Trueta 203, 08005, Barcelona, Spain (L.S.-G.); School of Earth, Atmospheric and Environmental Sciences, The University of Manchester, Oxford Road, Manchester M13 9PL, UK (B.E.v.D.); Norwegian Geotechnical Institute, Sognsveien 72, 0855, Oslo, Norway (V.A.).
*These authors contributed equally to this work.

subsidence, with ICD loss exacerbated by the increased wave and wind erosion that accompany sea-level rise and longer ice-free seasons[2]. Satellite images show a large erosional turbidity cloud along the ESAS coastline (Fig. 1b). From limited land-based surveys, this ICD erosion is thought to be delivering as much total organic carbon to the ESAS as all its large rivers combined[19,20]. Unfortunately, these studies are limited in spatial coverage, and do not consider the fate of the released carbon in the receiving ocean. There are no field-based reports of degradation or greenhouse-gas releases of thawing ICD; however, a recent investigation of organic matter genesis in ESS surface sediments suggests that ICD erosion may dominate over planktonic and riverine sources[21]. Laboratory experiments have shown that microbial degradation begins once permafrost has thawed, implying survival of viable bacteria and an inherent lability of the very old ICD organic carbon (ICD-OC)[10,11]. In addition to terrestrial ICD, the ESAS sediments (inundated by seawater during the early Holocene epoch) also host large Pleistocene deposits, presumably containing carbon in quantities similar to those in the upper-1-m soil pool[6,8]. These reservoirs are subject to active sea-floor thermal erosion[16,17], potentially releasing as much organic carbon as coastal erosion and rivers[20]. Overall, carbon released from thawing and eroding coastal permafrost may play a quantitatively important role in the Arctic carbon cycle.

To evaluate the role of the ICD and subsea permafrost carbon (hereafter jointly referred to as ICD-PF) in the contemporary ESAS carbon cycle, we adopted an inverse approach based on deducing the contribution of this ICD-PF to carbon accumulating on the entire ESAS shelf. We analysed more than 200 sediment samples (see Methods Summary), collected during ship-based expeditions spanning the ESAS (Supplementary Fig. 2, Supplementary Methods). We used a dual-carbon-isotope ($\delta^{13}$C and $\Delta^{14}$C) mixing model, solved with a Monte Carlo simulation strategy to account for endmember uncertainties, to deconvolve the relative contributions from ICD-PF, plankton detritus and a terrestrial/topsoil component. We then combined the fractional contribution from ICD-PF with the radiochronologically constrained sediment accumulation flux (Methods Summary and Supplementary Methods) to derive the shelf-wide re-burial flux of old carbon from permafrost.

We examined the fate of thawing ICD-OC in ambient conditions on coastal slopes of Muostakh, an island in the southeastern Laptev Sea that is disappearing as a result of erosion rates of up to $20\,\mathrm{m\,yr^{-1}}$ (refs 19,20,22; Fig. 1a). Bulk carbon contents, and molecular and isotopic compositions of ICD-OC, were assessed in conjunction with *in situ* $CO_2$ evasion fluxes (Supplementary Methods) to assess susceptibility of the organic carbon to degradation before delivery into coastal waters.

Radiocarbon ages of surface-sediment organic carbon ranged between 10,800 and 7,300 $^{14}$C yr (Fig. 2a shows $\Delta^{14}$C values; see also Supplementary Table 1) in the western ESS and the Dmitry Laptev Strait, regions dominated by coastal erosion (Fig. 1b). Organic-carbon radiocarbon ages were also old in the southern ESS and the Laptev Sea, ranging from 7,800 to 3,200 $^{14}$C yr. Lateral shelf transport times are likely to be much smaller than these measured $^{14}$C ages[23], implying significant supply of pre-aged carbon to these sediments. $\delta^{13}$C values varied, from −28.3 to −25.2‰ near the coast, to −24.8 to −21.2‰ on the outer ESAS (Fig. 2b; Supplementary Table 1). In contrast to other world-ocean shelf seas, where the sediment organic carbon originates from planktonic and riverine sources, coastline and sediment erosion represent significant sources of organic carbon to the ESAS. The relative contribution of the three sources was deduced from their carbon isotope fingerprints. In addition to a marine source, with $\delta^{13}\mathrm{C} = -24 \pm 3.0$‰ and $\Delta^{14}\mathrm{C} = 60 \pm 60$‰ (mean ± standard deviation (s.d.); Supplementary Methods, Supplementary Figs 4, 5), we distinguish between two terrestrial sources: ICD-PF organic carbon (coastal, inland, and subsea; formed before inundation), with $\delta^{13}\mathrm{C} = -26.3 \pm 0.67$‰ and $\Delta^{14}\mathrm{C} = -940 \pm 84$‰ (Supplementary Fig. 4, Supplementary Table 4), and topsoil permafrost (topsoil-PF) organic carbon (drained from vegetation debris and the thin, surficial, annual thaw layer of the continuous permafrost regions of northeast Siberia), with $\delta^{13}\mathrm{C} = -28.2 \pm 1.96$‰ and $\Delta^{14}\mathrm{C} = -126 \pm 54$‰ (Supplementary Fig. 4, Supplementary Table 3 and Supplementary Methods). The endmember source assignments are based on an extensive compilation of circum-arctic literature data, yielding statistically robust and distinctive values for the three endmembers, as further explained in the Supplementary Information (Supplementary Text; Supplementary Figs 4, 5; Supplementary Tables 3, 4). Naturally, the isotopic endmember values carry uncertainties, which may be reduced in the future by additional observations of the marine and topsoil composition. The $^{13}$C and $^{14}$C compositions of the three endmembers are well separated from each other (Supplementary Fig. 4), which allows separation of their contributions while properly accounting for the associated uncertainties using the Monte Carlo simulation approach. We stress that the two terrestrial endmembers are solely source-based, and independent of transport or mobilization route, meaning that both ICD-PF and topsoil-PF can be delivered by coastal, delta and riverbank erosion as well as river transport. The resulting isotopic mass-balance model shows contributions of marine (planktonic) organic carbon to the shelf sediments ranging between 7% nearshore and 54% on the outer shelf, whereas topsoil-PF contributes ~30–35% close to land, decreasing to ~5% farther out (Fig. 2c).

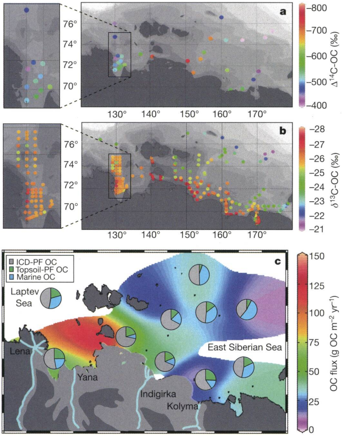


**Figure 2 | Carbon isotope compositions and contribution of organic carbon sources to sediment accumulation on the East Siberian Arctic Shelf.** **a**, **b**, $\Delta^{14}$C-OC (**a**) and $\delta^{13}$C-OC (**b**) signals in ESAS surface sediments. **c**, Annual sedimentary organic carbon accumulation fluxes ($\mathrm{g\,OC\,m^{-2}\,yr^{-1}}$) and relative contributions (pie charts) of the three source pools to the surface-sediment organic carbon on the ESAS. The mean ESS contributions are: 57 ± 1.6% from ICD-PF (grey), 16 ± 3.4% from topsoil-PF (green) and 26 ± 8.0% from marine/planktonic organic carbon (blue), as identified by numerical (Monte Carlo) simulations of the dual-carbon-isotope ($\delta^{13}$C and $\Delta^{14}$C) and endmember mixing models. Land area marked in light grey indicates the distribution of the Ice Complex[30].

ICD-PF constitutes 36–76% of the sedimentary organic carbon throughout the broad shelf, despite its largely coastal delivery. ICD-OC is ballasted by mineral association and rapidly settles[21,24], whereupon it is probably resuspended from the sea floor and dispersed over the shelf, mostly by bottom-boundary-layer transport[21,25,26]. Old permafrost-released erosional carbon thus dominates burial of organic carbon on the ESAS.

We estimate the net sediment burial of ICD-PF carbon using accumulation fluxes from sediment cores ($36 \pm 17$ g OC m$^{-2}$ yr$^{-1}$; all confidence intervals are 95%, unless otherwise stated; Fig. 2c, Supplementary Table 2). This was scaled up by the fraction of sea floor that is available for carbon burial (0.6), corresponding to water depth $>30$ m (Supplementary Fig. 2), where resuspension is negligible and sediments thus accumulate[26]. Combining the ESS shelf area ($9.87 \times 10^5$ km$^2$) with the ICD-PF contribution to the sediment organic carbon (ESS only: $57 \pm 1.6\%$; Supplementary Table 5) yields an overall annual ICD-PF carbon accumulation flux of $12 \pm 8$ TgC yr$^{-1}$. Inclusion of the Laptev Sea increases this value to $20 \pm 8$ TgC yr$^{-1}$ (Supplementary Table 6). Hence, this approach reveals that the supply of carbon from ICD-PF erosion to the ESAS is much larger than has previously been assumed[14,19,20].

The biogeochemical composition of the eroding slopes of Muostakh Island (Fig. 3) indicates extensive organic matter degradation of the thawing ICD before delivery to the ocean. Recurring trends were observed in several properties between higher and lower elevations on the investigated slopes that are consistent with continuing degradation (Fig. 3; Supplementary Tables 7, 8), specifically: decreasing soil organic carbon content; increasing $\delta^{13}$C of organic carbon ($\delta^{13}C_{OC}$); decreasing $\Delta^{14}C_{OC}$; decreasing ratio of high-molecular-weight *n*-alkanoic acids to high-molecular-weight *n*-alkanes; increasing ratio of even, low-molecular-weight to odd, high-molecular-weight *n*-alkanes; and increase in atmospheric CO$_2$ venting, deduced from field-chamber soil respiration measurements (Supplementary Methods). These trends and fluxes contrast with prior assumptions that all thawed and erosion-mobilized ICD-OC is directly flushed into the sea without sub-aerial degradation[14,19,20]. The elemental, isotopic and molecular data imply $66 \pm 16\%$ (mean ± s.d.; Supplementary Methods) down-slope degradative loss of ICD-OC.

Combining the $20 \pm 8$ Tg C yr$^{-1}$ sediment re-burial flux of thawed old organic carbon with a recent estimate of water-column degradation of terrestrially derived particulate organic carbon on the ESAS of 1.4 yr$^{-1}$ ($2.5 \pm 1.6$ Tg C yr$^{-1}$; mean ± s.d.)[27] suggests an ICD-PF organic carbon flux to the marine system of $22 \pm 8$ Tg C yr$^{-1}$ (Supplementary Fig. 1). Assuming an equal contribution of this flux from coastline and subsea erosion (Supplementary Table 6, which also includes 25/75% and 75/25% models), the $66 \pm 16\%$ carbon loss along the eroding coastal slopes corresponds to a carbon venting (presumably mostly CO$_2$) from the ICD of $22 \pm 8$ Tg yr$^{-1}$ (Supplementary Fig. 1). The total remobilization of old organic carbon from thawing of ICD-PF is thus $\sim 44 \pm 10$ Tg C yr$^{-1}$ (Supplementary Table 6; Supplementary Fig. 1).

The present assessment suggests a substantially larger flux of carbon from thawing ICD permafrost ($44 \pm 10$ Tg C yr$^{-1}$; Supplementary Table 6) than has been inferred previously from exclusively land-based surveys ($\sim 4$ Tg C yr$^{-1}$; no error reported)[14]. Previous estimates of ICD erosion may have been too low for several reasons, including gross upscaling from limited point measurements of ICD retreat rates[19,20,22]. In addition, upscaling using digital shoreline length data leads to considerable underestimations[28]; and potentially large inputs from retrogressive thaw slumps and slope failure[28] are excluded when elevation change data are not included in coastline retreat measurements. Finally, bottom erosion is a previously neglected but potentially important contributor of old eroded organic carbon to the modern biogeochemical cycle on the ESAS, with erosion rates of 10–30 cm yr$^{-1}$ (refs 18,29) at depths less than 30 m (nearly half the ESAS), where present-day bottom-water temperatures in summer are 2–3 °C and

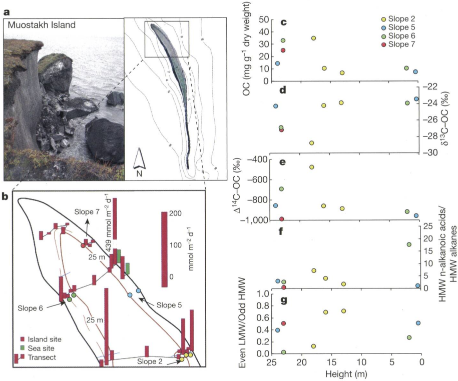


**Figure 3 | Biogeochemical signals of Ice Complex organic matter degradation on Muostakh Island.** **a**, Study area. **b**, Distribution of CO$_2$ outgassing. **c–g**, Distributions along the four studied slopes (positions indicated in **b**) of soil organic carbon content (**c**); $\delta^{13}$C-OC signal (**d**); $\Delta^{14}$C-OC signal (**e**); ratio of high-molecular-weight *n*-alkanoic acids to high-molecular-weight *n*-alkanes (proxy for degradation status) (**f**) and ratio of even, low-molecular-weight *n*-alkanes to odd, high-molecular-weight *n*-alkanes (proxy for bacterial biomass relative to substrate) (**g**). Ratios in **f** and **g** are molecular ratios.

have risen during the past decade[13]. Thermal collapse of the carbon-rich, permafrost-laden coastlines and sea floors may accelerate with Arctic amplification of climate warming, and could further intensify the role of old Ice Complex organic carbon in carbon cycling in the world's largest shelf sea.

## METHODS SUMMARY

Surface sediments were collected on several expeditions on the ESAS in 2004, 2005, 2007 and 2008 (Supplementary Fig. 2, Supplementary Tables 1 and 9). The samples were analysed for organic carbon content and $\delta^{13}$C (UC Davis Stable Isotope Facility, USA) and $\Delta^{14}$C (US National Ocean Sciences Accelerator Mass Spectrometry (NOSAMS) Facility of the Woods Hole Oceanographic Institution, USA). The relative contributions of three endmember sources—Coastal Ice Complex permafrost (ICD-PF: $\delta^{13}\mathrm{C} = -26.3 \pm 0.67$‰; $\Delta^{14}\mathrm{C} = -940 \pm 84$‰; Supplementary Table 4); topsoil permafrost (topsoil-PF: $\delta^{13}\mathrm{C} = -28.2 \pm 1.96$‰; $\Delta^{14}\mathrm{C} = -126 \pm 54$‰; Supplementary Table 3); and marine organic carbon ($\delta^{13}\mathrm{C} = -24 \pm 3.0$‰, $\Delta^{14}\mathrm{C} = 60 \pm 60$‰; Supplementary Figs 4, 5)—to the surface sediment organic carbon content were quantified using a dual-carbon-isotope mixing model, solved with a Monte Carlo simulation approach (Supplementary Table 3). Radiochronological measurements on sediment cores from the ESAS were performed at Stockholm University and at the Radiation Research Division of the Risø National Laboratory for Sustainable Energy, Denmark (Supplementary Table 10, Supplementary Fig. 3). Total inventories of excess $^{210}$Pb were used to calculate the annual sediment organic carbon accumulation on the ESAS (Supplementary Table 2). The average contribution of organic carbon from ICD-PF in the surface sediment was then used to infer the annual sediment organic carbon accumulation from ICD-PF to the ESAS.

Ice Complex samples from the slopes of Muostakh Island were collected in July 2006 (Fig. 3, Supplementary Table 7). Bulk organic carbon and $\delta^{13}$C analyses were performed at Stockholm University (Department of Geological Sciences) and $\Delta^{14}$C analyses at NOSAMS. The soil samples were extracted and separated for identification of molecular biomarkers using gas chromatography/mass spectrometry. In addition, soil respiration measurements were collected on Muostakh Island slopes with automatic lid chambers equipped with infrared gas analysers (Fig. 3; Supplementary Table 8). Full details of methods are available in Supplementary Methods.

**Received 1 December 2011; accepted 3 July 2012.**

**Published online 29 August 2012.**

1. Friedlingstein, P. *et al.* Climate-carbon cycle feedback analysis: results from the C$^4$MIP model intercomparison. *J. Clim.* **19,** 3337–3353 (2006).
2. Solomon, S. D, *et al.* (eds) *Climate Change 2007: The Physical Science Basis* (Cambridge Univ. Press, 2007).
3. Gruber, N. *et al.* in *The Global Carbon Cycle: Integrating Humans, Climate and the Natural World*, (eds Field, C. B. & Raupach, M. R.) 45–76 (Island Press, 2004).
4. Zimov, S. A., Schuur, E. A. G. & Chapin, F. S. III. Permafrost and the global carbon budget. *Science* **312,** 1612–1613 (2006).
5. Schirrmeister, L. *et al.* Sedimentary characteristics and origin of the Late Pleistocene Ice Complex on north-east Siberian Arctic coastal lowlands and islands – a review. *Quat. Int.* **241,** 3–25 (2011).
6. Soloviev, V. A., Ginzburg, G. D., Telepnev, E. V. & Mikhaluk, Y. N. *Cryothermia of Gas Hydrates in the Arctic Ocean* (VNIIOkeangeologia, 1987).
7. Shakhova, N. *et al.* Extensive methane venting to the atmosphere from sediments of the East Siberian Arctic Shelf. *Science* **327,** 1246–1250 (2010).
8. Tarnocai, C. *et al.* Soil organic carbon pools in the northern circumpolar permafrost region. *Glob. Biogeochem. Cycles* **23,** GB2023 (2009).
9. Schuur, E. A. G. *et al.* The effect of permafrost thaw on old carbon release and net carbon exchange from tundra. *Nature* **459,** 556–559 (2009).
10. Rivkina, E., Gilichinsky, D., Wagener, S., Tiedje, J. & McGrath, J. Biogeochemical activity of anaerobic microorganisms from buried permafrost sediments. *Geomicrobiol. J.* **15,** 187–193 (1998).
11. Dutta, K., Schuur, E. A. G., Neff, J. C. & Zimov, S. A. Potential carbon release from permafrost soils of Northeastern Siberia. *Glob. Change Biol.* **12,** 2336–2351 (2006).
12. Richter-Menge, J. & Overland, J. E. (eds) *Arctic Report Card 2010,* http://www.arctic.noaa.gov/reportcard (2010).
13. Dmitrenko, I. A. *et al.* Recent changes in shelf hydrography in the Siberian Arctic: potential for subsea permafrost instability. *J. Geophys. Res.* **116,** C10027 (2011).
14. Stein, R. & Macdonald, R. W. *The Organic Carbon Cycle in the Arctic Ocean* (Springer, 2004).
15. Mastepanov, M. *et al.* Large tundra methane burst during onset of freezing. *Nature* **456,** 628–630 (2008).
16. Nicolsky, D. & Shakhova, N. Modeling sub-sea permafrost in the East Siberian Arctic Shelf: the Dmitry Laptev Strait. *Environ. Res. Lett.* **5,** 015006 (2010).
17. Romanovskii, N. N., Hubberten, H.-W., Gavrilov, A. V., Eliseeva, A. A. & Tipenko, G. S. Offshore permafrost and gas hydrate stability zone on the shelf of East Siberian Seas. *Geo-Mar. Lett.* **25,** 167–182 (2005).
18. Grigoriev, M. N. *Cryomorphogenesis and Lithodynamics of the Coastal-shelf Zone of the Seas of Eastern Siberia.* Doctoral thesis, Yakutsk Melnikov Permafrost Inst. (2008).
19. Semiletov, I. P. The failure of coastal frozen rock as an important factor in the biogeochemistry of the Arctic shelf water. *Dokl. Earth Sci.* **369,** 1140–1143 (1999).
20. Rachold, V. *et al.* Coastal erosion vs riverine sediment discharge in the Arctic Shelf seas. *Int. J. Earth Sci.* **89,** 450–460 (2000).
21. Vonk, J. E. *et al.* Molecular and radiocarbon constraints on sources and degradation of terrestrial organic carbon along the Kolyma paleoriver transect, East Siberian Sea. *Biogeosciences* **7,** 3153–3166 (2010).
22. Overduin, P. P. *et al.* The evolution and degradation of coastal and offshore permafrost in the Laptev and East Siberian Seas during the last climatic cycle. *Geol. Soc. Am. Spec. Pap.* **426,** 97–110 (2007).
23. Keil, R. G., Dickens, A. F., Arnarson, T., Nunn, B. L. & Devol, A. H. What is the oxygen exposure time of laterally transported organic matter along the Washington margin? *Mar. Chem.* **92,** 157–165 (2004).
24. Vonk, J. E., van Dongen, B. E. & Gustafsson, Ö. Selective preservation of old organic carbon fluvially released from sub-Arctic soils. *Geophys. Res. Lett.* **37,** L11605 (2010).
25. Wegner, C. *et al.* Suspended particulate matter on the Laptev Sea shelf (Siberian Arctic) during ice-free conditions. *Estuar. Coast. Shelf Sci.* **57,** 55–64 (2003).
26. Dudarev, O. V., Semiletov, I. P., Charkin, A. N. & Botsul, A. I. Deposition settings on the continental shelf of the East Siberian Sea. *Dokl. Earth Sci.* **409,** 1000–1005 (2006).
27. Sánchez-García, L. *et al.* Inventories and behavior of particulate organic carbon in the Laptev and East Siberian seas. *Glob. Biogeochem. Cycles* **25,** GB2007 (2011).
28. Lantuit, H. *et al.* Towards a calculation of organic carbon release from erosion of Arctic coasts using non-fractal coastline datasets. *Mar. Geol.* **257,** 1–10 (2009).
29. Razumov, S. O. Rates of coastal thermoabrasion as a function of climate and morphological characteristics of the coast. *Geomorphology* **3,** 88–94 (2000).
30. Romanovskii, N. N. *Fundamentals of the Cryogenesis of the Lithosphere* (Moscow University Press, Moscow, 1993).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** We thank all ISSS-08 colleagues and crew, in particular M. Kruså, P. Andersson and V. Mordukhovich, who helped with sampling. The ISSS program is supported by the Knut and Alice Wallenberg Foundation, the Far Eastern Branch of the Russian Academy of Sciences, the Swedish Research Council, the US National Oceanic and Atmospheric Administration, the Russian Foundation of Basic Research, the Swedish Polar Research Secretariat and the Nordic Council of Ministers (Arctic Co-Op and TRI-DEFROST programs). Ö.G. and L.S.-G. acknowledge an Academy Research Fellow grant from the Swedish Royal Academy of Sciences and an EU Marie Curie grant, respectively. N.S. and I.P.S. acknowledge grants from the US National Science Foundation and the NOAA OAR Climate Program Office.

**Author Contributions** All authors except P.R., T.I.E. and A.A. collected samples. Preparations for bulk organic carbon analyses, stable isotope analysis and radiocarbon analyses were made by J.E.V. (sediments) and L.S.-G. (Ice Complex samples). Radiocarbon analyses on sediments were facilitated by T.I.E. L.S.-G. analysed lipid biomarkers in Muostakh Island samples. A.A. was responsible for the Monte Carlo simulations. Radiochronological measurements on sediment cores were made by P.R. and at Stockholm University. J.E.V. performed data analyses and flux calculations. J.E.V., L.S.-G. and Ö.G. wrote the paper, with input from N.S., I.P.S. and all other authors.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Ö.G. (orjan.gustafsson@itm.su.se).

doi:10.1038/nature11391

# Recent Antarctic Peninsula warming relative to Holocene climate and ice-shelf history

Robert Mulvaney[1], Nerilie J. Abram[1,2], Richard C. A. Hindmarsh[1], Carol Arrowsmith[3], Louise Fleet[1], Jack Triest[1], Louise C. Sime[1], Olivier Alemany[4] & Susan Foord[1]‡

**Rapid warming over the past 50 years on the Antarctic Peninsula is associated with the collapse of a number of ice shelves and accelerating glacier mass loss[1–7]. In contrast, warming has been comparatively modest over West Antarctica and significant changes have not been observed over most of East Antarctica[8,9], suggesting that the ice-core palaeoclimate records available from these areas may not be representative of the climate history of the Antarctic Peninsula. Here we show that the Antarctic Peninsula experienced an early-Holocene warm period followed by stable temperatures, from about 9,200 to 2,500 years ago, that were similar to modern-day levels. Our temperature estimates are based on an ice-core record of deuterium variations from James Ross Island, off the northeastern tip of the Antarctic Peninsula. We find that the late-Holocene development of ice shelves near James Ross Island was coincident with pronounced cooling from 2,500 to 600 years ago. This cooling was part of a millennial-scale climate excursion with opposing anomalies on the eastern and western sides of the Antarctic Peninsula. Although warming of the northeastern Antarctic Peninsula began around 600 years ago, the high rate of warming over the past century is unusual (but not unprecedented) in the context of natural climate variability over the past two millennia. The connection shown here between past temperature and ice-shelf stability suggests that warming for several centuries rendered ice shelves on the northeastern Antarctic Peninsula vulnerable to collapse. Continued warming to temperatures that now exceed the stable conditions of most of the Holocene epoch is likely to cause ice-shelf instability to encroach farther southward along the Antarctic Peninsula.**

The Antarctic Peninsula is at present one of the most rapidly warming regions on Earth[1] (Fig. 1a). Historical observations since 1958 at Esperanza Station (Fig. 1b) document warming equivalent to 3.5 ± 0.8 °C per century. During this time, a series of ice shelves stretching from Prince Gustav Channel to the Larsen B ice shelf on the northeastern Antarctic Peninsula have been lost[2–5], causing an acceleration of the feeder glaciers that drain ice from the Antarctic Peninsula[6]. To assess these recent rapid changes, a longer-term perspective on Antarctic Peninsula climate and the role of past atmospheric temperature in determining ice-shelf stability is urgently needed[7]. To address this, we drilled an ice core to the bed of the ice cap on James Ross Island (JRI). This site lies off the northeastern tip of the Antarctic Peninsula, adjacent to the area that has witnessed a series of ice-shelf collapses since 1995 (Fig. 1b).

The 363.9-m-long JRI ice core provides a temperature reconstruction, based on deuterium/hydrogen isotope ratios of the ice (δD), that spans the entire Holocene and extends into the last glacial interval (Fig. 2, Methods Summary and Supplementary Fig. 1). Evidence of the glacial age ice is found in the final 5 m of the JRI ice core; initial estimates suggest the record may extend to ~50,000 yr BP (by convention, 0 yr BP means AD 1950), although an unrealistically rapid isotopic transition implies that an unconformity may be present in the early deglacial interval of the ice core. Taking into account changes in ocean isotopic values[10,11], the isotopic composition of the glacial ice on JRI is equivalent to temperatures that were approximately 6.1 ± 1.0 °C

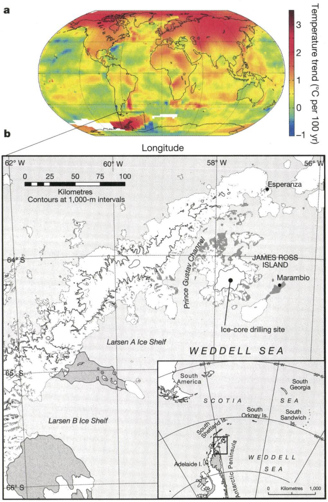


**Figure 1 | Regional and climatic setting of the Antarctic Peninsula.** **a**, Temperature trends for the 50 years from 1958 to 2008 show the rapid regional warming of the Antarctic Peninsula. Trends are shown for January–December annual averages of gridded land and ocean surface temperature data[27,28]. **b**, James Ross Island (JRI) is located near the northeastern tip of the Antarctic Peninsula, within the zone of rapid regional warming, and adjacent to the former Prince Gustav, Larsen A and Larsen B ice shelves.

---

[1]British Antarctic Survey, Natural Environment Research Council, Cambridge CB3 0ET, UK. [2]Research School of Earth Sciences, The Australian National University, Canberra, Australian Capital Territory 0200, Australia. [3]NERC Isotope Geosciences Laboratory, Keyworth NG12 5GG, UK. [4]UJF – Grenoble 1/CNRS, Laboratoire de Glaciologie et Géophysique de l'Environnement (LGGE) UMR 5183, Grenoble F-38041, France.
‡Deceased.

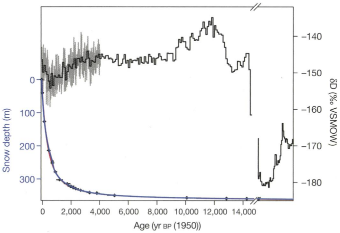


**Figure 2 | Isotope and depth–age profiles of the JRI ice core.** The δD isotope profile for the JRI ice core is shown in terms of the 100-yr average (black) for the whole of the Holocene and the 10-yr average (grey) for 4,000 yr BP to present. The JRI depth–age scale, JRI-1 (blue), was constructed on the basis of a glaciological flow model for this site (red) with adjustment derived from fixed time markers (black diamonds; horizontal error bars give estimated age uncertainty for the fixed markers). Further details are provided in Methods and Supplementary Table 1.

cooler than present (where by present we mean AD 1961–1990) during the Last Glacial Maximum[12,13] (LGM). By comparison, the LGM is found to have been 7.4 °C cooler in Dronning Maud Land and 9.3 °C cooler at Dome C on the East Antarctic plateau[14].

The reduced magnitude of LGM-Holocene temperature change on the Antarctic Peninsula probably reflects its more northerly position and proximal maritime influence. An alternative explanation could be that the JRI ice cap experienced changes in elevation at the LGM, making this site seem isotopically warmer than continental Antarctica. However, this interpretation would require that the JRI ice cap at the LGM was ~150–360 m lower than present[13], according to Dronning Maud Land and Dome C temperatures[14]. Such a reduction is inconsistent with glaciological evidence that the JRI ice cap had a confluence with the Antarctic Peninsula ice sheet in the Prince Gustav Channel until the early Holocene[15]. The JRI ice core thus adds to the glaciological history of the northern Antarctic Peninsula, with the reduced LGM-Holocene isotope contrast implying that the ice cap cannot have thickened significantly at the LGM and was not overrun by isotopically colder ice from the south.

The Holocene temperature history from the JRI ice core is characterized by an early-Holocene climatic optimum that was 1.3 ± 0.3 °C warmer than present (Fig. 3). The magnitude and progression of this early-Holocene optimum is similar to that observed in ice-core records from the main Antarctic continent[16]. A marine sediment record from off the shore of the western Antarctic Peninsula also shows an early-Holocene optimum during which surface ocean temperatures were determined to be ~3.5 °C higher than present[17]. Other evidence suggests that the George VI ice shelf on the southwestern Antarctic Peninsula was absent during this early-Holocene warm interval but reformed in the mid Holocene[7].

Following this widespread early-Holocene climate optimum, temperature on the Antarctic Peninsula decreased and the JRI ice core documents a long interval of stable climate that persisted from ~9,200 to 2,500 yr BP (Fig. 3). During this interval, the mean temperature anomaly, of 0.2 ± 0.2 °C, indicates that conditions at JRI were comparable to the warm conditions observed at this site over recent decades. Likewise, marine temperatures on the western side of the Antarctic Peninsula[17] declined to reach, by ~8,000 yr BP, a long-term mean that was close to present-day values. Within this interval of mid-Holocene stability, the JRI isotope record indicates that from ~5,000 to 3,000 yr BP conditions may have been only marginally warmer than present. Various proxy evidence exists for a mid-Holocene warm period on the Antarctic Peninsula[7], although the lack of a consensus on its timing in this region may be explained by the small magnitude of this feature in the JRI temperature record compared with the well-defined mid-Holocene climate optimum in continental Antarctic ice-core records[16].

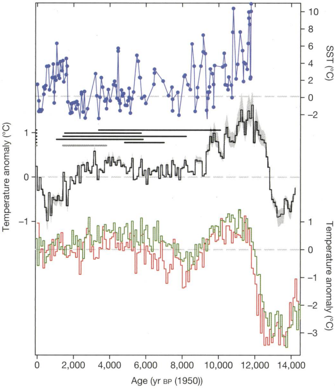


**Figure 3 | Holocene temperature history of the Antarctic Peninsula.** The JRI ice-core temperature reconstruction relative to the 1961–1990 mean (black trace, 100-yr average; the grey band indicates the standard error of the calibration dependence) is shown alongside a sea surface temperature (SST) reconstruction from off the shore of the western Antarctic Peninsula (blue curve)[17], and temperature reconstructions from the Dome C (red)[29] and Dronning Maud Land (green)[30] ice cores from East Antarctica. Horizontal bars show intervals in the Holocene when marine sediment cores indicate that open water was present in the area of the Prince Gustav (black; top to bottom are north to south core sites; original $^{14}$C ages have been calibrated)[3] and Larsen A (grey)[5] ice shelves, which collapsed in AD 1995.

The Holocene ice-shelf history along the eastern Antarctic Peninsula shows a strong connection to Antarctic Peninsula temperatures. Following the deglacial transition from grounded to floating ice in Prince Gustav Channel at ~10,000 to 8,000 yr BP[3,15], this area experienced intervals of seasonally open water through to ~1,500 yr BP[3]. Marine sediments indicate that a permanent ice shelf was established there only after ~1,500 yr BP and that the maximum ice-shelf extent may have been reached as recently as a few centuries ago[3]. Farther south, there is evidence for instability of the Larsen A ice shelf between 3,800 and 1,400 yr BP[5]. Farther south again, the Larsen B ice shelf probably remained intact throughout the Holocene, although there is evidence that the ice shelf was progressively weakened by melting[4]. Combining the JRI temperature reconstruction with the marine sediment evidence shows that temperatures similar to present occurred in this region for much of the Holocene, resulting in a regime in which ice shelves were only transient features along the northern-most part of the eastern Antarctic Peninsula and were undergoing decay farther to the south. An additional new perspective is that recent warming to levels consistent with the mid Holocene meant that the ice shelves along the northeastern Peninsula were poised for the succession of collapses observed there over recent decades.

The late-Holocene development of ice shelves fed from the northeastern Antarctic Peninsula seems to be related to millennial-scale climate variability in the region (Figs 3 and 4a). After 2,500 yr BP, the JRI isotope record documents pronounced cooling to temperatures that were on average 0.7 ± 0.3 °C cooler than present between 800 and 400 yr BP (AD 1150–1550), and on a decadal timescale temperatures may have at times been more than 1.8 ± 0.3 °C cooler than present. Late-Holocene cooling has also been inferred from northeastern Antarctic Peninsula lake records[7,18]. The prominent millennial-scale cooling at JRI is matched by a similarly prominent but warm excursion in marine temperatures to the west of the Antarctic Peninsula[17,19]. On the central spine of the Antarctic Peninsula, a 500-yr-long ice-core record from the Dyer Plateau shows that temperatures here were approximately the same as present at 450 yr BP[20], suggesting an east–west divide across the Antarctic Peninsula in this late-Holocene climate oscillation. Thus, although glacial-scale climate changes have been consistent across the whole of the Antarctic Peninsula region, millennial-scale climate variability was particularly strong during the late Holocene and seems to have been characterized by opposing east–west temperature anomalies across the Antarctic Peninsula.

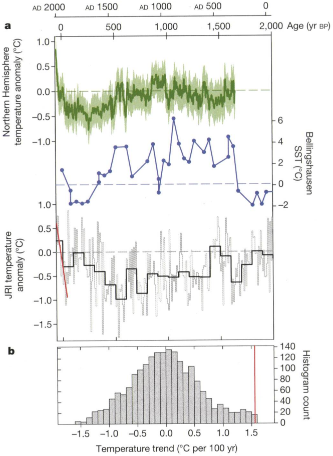


**Figure 4 | Two-thousand-year climate history of the Antarctic Peninsula.** **a**, The JRI temperature reconstruction (black trace, 100-yr average; grey trace, 10-yr average; relative to 1961–1990 mean) is shown alongside the SST record from Ocean Drilling Program site 1098 to the west of the Antarctic Peninsula[17] (blue curve) and the reconstructed Northern Hemisphere temperature anomaly[22] (dark green curve, relative to 1961–1990 mean; the light green envelope indicates the 95% confidence interval). Whereas SST to the west of the Antarctic Peninsula shows similarities to Northern Hemisphere climate over the past 2,000 yr, the JRI record shows an opposing temperature excursion which demonstrates that the Antarctic Peninsula did not experience a widespread Medieval Warm Period/Little Ice Age sequence comparable to Northern Hemisphere climate at that time. Warming at JRI has been ongoing for several centuries, although the warming by 1.56 °C over the past 100 yr (red lines in **a** and **b**) is highly unusual in the context of natural variability. **b**, This is shown by a histogram analysis of temperature trends calculated in moving 100-yr windows of annual-resolution data from the JRI ice core starting at 2,000 yr BP.

Opposing temperature anomalies on either side of the Antarctic Peninsula are a feature of the Antarctic dipole, which is an interannual standing-wave pattern that results in opposite temperature and sea ice anomalies between the Weddell Sea and the Amundsen and Bellingshausen seas[21]. The observation of similar opposing climate oscillations on a millennial scale provides an indication that the Antarctic dipole may also influence long-term climate changes in the Antarctic Peninsula region. Deducing the exact mechanisms that have driven this late-Holocene Antarctic-dipole-like pattern will require additional, well-dated palaeoclimate reconstructions to map the spatial extent of the climate anomalies. We note, however, that the development of this Antarctic-dipole-like feature during the late Holocene coincides with the well-documented maximum in El Niño activity (Supplementary Fig. 2), which has a role in driving present-day variability of the Antarctic dipole[21]. Antarctic-dipole-like cooling of the Weddell Sea in the late Holocene, and the propagation of these ocean temperature and sea ice anomalies along the eastern Antarctic Peninsula by the Weddell gyre, may have also aided the rapid establishment of ice shelves in this region during the late Holocene.

Sustained warming at JRI began ~600 yr ago (Fig. 4a). Lake sediments from Beak Island in Prince Gustav Channel also indicate warming beginning at ~AD 1410[18], and together these records demonstrate the absence of a widespread Little Ice Age signal on the Antarctic Peninsula that was comparable to Northern Hemisphere climate[22] (Fig. 4a). The overall rate of pre-anthropogenic temperature increase at JRI from AD 1400 to AD 1850 equates to 0.22 ± 0.06 °C per century. However, there are times in this interval when warming occurred much faster. Using annual-resolution data, trends were calculated for the JRI temperature record since 2,000 yr BP over moving 100-yr intervals stepped in 1-year increments (yielding 1,958 100-year analysis windows) (Fig. 4b). This analysis indicates that rapid warming trends exceeding 1.5 °C per century occurred at JRI during the intervals spanning AD 1518–1621 and AD 1671–1777, and that trends exceeding 1.25 °C per century occurred during the interval AD 296–415.

Over the past 100 yr, the JRI ice-core record shows that the mean temperature there has increased by 1.56 ± 0.42 °C (Fig. 4a). This ranks as one of the fastest (upper 0.3%) warming trends at JRI since 2,000 yr BP, according to the moving 100-yr analysis windows, demonstrating that rapid recent warming of the Antarctic Peninsula is highly unusual although not outside the bounds of natural variability in the pre-anthropogenic era (Fig. 4b). The JRI ice core shows that the recent phase of warming on the northern Antarctic Peninsula began in the mid 1920s and that over the past 50 yr the temperature has risen at a rate equivalent to 2.6 ± 1.2 °C per century. Repeating the temperature trend analysis using 50-yr windows confirms the finding that the rapidity of recent Antarctic Peninsula warming is unusual but not unprecedented.

The long-term climate history provided by the JRI ice core shows that natural millennial-scale climate variability has resulted in warming on the eastern Antarctic Peninsula that has been ongoing for a number of centuries and had left ice shelves in this area vulnerable to collapse during the recent phase of rapid warming. If warming continues in this region, as is suggested by its attribution in part to rising atmospheric greenhouse gas concentrations[7,23], then temperatures will soon exceed the stable conditions that persisted in the eastern Antarctic Peninsula for most of the Holocene. The association between atmospheric temperature and ice-shelf stability in the past demonstrates that as warming continues ice-shelf vulnerability is likely to progress farther southwards along the Antarctic Peninsula coast to affect ice shelves that have been stable throughout the Holocene, and may make them particularly susceptible to changes in oceanographic forcing[24].

## METHODS SUMMARY

The JRI ice core discussed in this study was drilled in January–February 2008 at a site (57° 41.10′ W, 64° 12.10′ S, 1,542-m elevation; Fig. 1) near the summit of Mount Haddington. The ice core was recovered to bedrock at a depth of 363.9 m. The mean annual temperature at this site is −14.4 °C, and the mean annual snow accumulation is 0.63 m water equivalent[12,25]. The Holocene age scale for the JRI ice core, termed JRI-1, is based on a glaciological flow model with additional age control provided by fixed time markers derived from local and global volcanic events (Fig. 2 and Supplementary Table 1). The temperature reconstruction was based on deuterium isotope (δD) measurements (expressed relative to the international standard Vienna Standard Mean Ocean Water (VSMOW) along the length of the ice core, with a typical precision of 1.0‰. Temperature anomalies were calculated using a δD–temperature dependence of 6.4 ± 1.3‰ °C$^{-1}$ (ref. 12), under the assumption that the modern-day calibration holds over the entire record[26], and are given with reference to AD 1961–1990. Consistent palaeotemperature results are produced using the oxygen isotopic ratio (δ$^{18}$O) of the ice (Supplementary Fig. 1), confirming that the isotopic record primarily reflects changes in temperature at the JRI site during the Holocene. Uncertainties in mean temperature anomalies are the combined standard error of the calibration dependence and standard deviation of the variability of 100-yr-binned data.

**Full Methods** and any associated references are available in the online version of the paper.

**Received 11 November 2011; accepted 29 June 2012.**

**Published online 22 August 2012.**

1. Vaughan, D. G. *et al.* Recent rapid regional climate warming on the Antarctic Peninsula. *Clim. Change* **60,** 243–274 (2003).
2. Pudsey, C. J. & Evans, J. First survey of Antarctic sub-ice shelf sediments reveals mid-Holocene ice shelf retreat. *Geology* **29,** 787–790 (2001).
3. Pudsey, C. J., Murray, J. W., Appleby, P. & Evans, J. Ice shelf history from petrographic and foraminiferal evidence, Northeast Antarctic Peninsula. *Quat. Sci. Rev.* **25,** 2357–2379 (2006).
4. Domack, E. *et al.* Stability of the Larsen B ice shelf on the Antarctic Peninsula during the Holocene epoch. *Nature* **436,** 681–685 (2005).
5. Brachfeld, S. *et al.* Holocene history of the Larsen-A Ice Shelf constrained by geomagnetic paleointensity dating. *Geology* **31,** 749–752 (2003).
6. Cook, A. J., Fox, A. J., Vaughan, D. G. & Ferrigno, J. G. Retreating glacier fronts on the Antarctic Peninsula over the past half-century. *Science* **308,** 541–544 (2005).
7. Bentley, M. J. *et al.* Mechanisms of Holocene palaeoenvironmental change in the Antarctic Peninsula region. *Holocene* **19,** 51–69 (2009).
8. Turner, J. *et al.* Antarctic climate change during the last 50 years. *Int. J. Climatol.* **25,** 279–294 (2005).
9. Steig, E. J. *et al.* Warming of the Antarctic ice-sheet surface since the 1957 International Geophysical Year. *Nature* **457,** 459–462 (2009).
10. Jouzel, J. *et al.* Magnitude of isotope/temperature scaling for interpretation of central Antarctic ice cores. *J. Geophys. Res.* **108,** 4361 (2003).
11. Bintanja, R., van de Wal, R. S. W. & Oerlemans, J. Modelled atmospheric temperatures and global sea levels over the past million years. *Nature* **437,** 125–128 (2005).
12. Abram, N. J., Mulvaney, R. & Arrowsmith, C. Environmental signals in a highly resolved ice core from James Ross Island, Antarctica. *J. Geophys. Res.* **116,** D20116 (2011).
13. Masson-Delmotte, V. *et al.* A review of Antarctic surface snow isotopic composition: observations, atmospheric circulation, and isotopic modeling. *J. Clim.* **21,** 3359–3387 (2008).
14. Stenni, B. *et al.* The deuterium excess records of EPICA Dome C and Dronning Maud Land ice cores (East Antarctica). *Quat. Sci. Rev.* **29,** 146–159 (2010).
15. Johnson, J. S., Bentley, M. J., Roberts, S. J., Binnie, S. A. & Freeman, S. P. H. T. Holocene deglacial history of the northeast Antarctic Peninsula: a review and new chronological constraints. *Quat. Sci. Rev.* **30,** 3791–3802 (2011).
16. Masson-Delmotte, V. *et al.* A comparison of the present and last interglacial periods in six Antarctic ice cores. *Clim. Past* **7,** 397–423 (2011).
17. Shevenell, A. E., Ingalls, A. E., Domack, E. W. & Kelly, C. Holocene Southern Ocean surface temperature variability west of the Antarctic Peninsula. *Nature* **470,** 250–254 (2011).
18. Sterken, M. *et al.* Holocene glacial and climate history of Prince Gustav Channel, northeastern Antarctic Peninsula. *Quat. Sci. Rev.* **31,** 93–111 (2012).
19. Hall, B. L., Koffman, T. & Denton, G. H. Reduced ice extent on the western Antarctic Peninsula at 700–970 cal. yr BP. *Geology* **38,** 635–638 (2010).
20. Thompson, L. G. *et al.* Climate since 1520 AD on Dyer Plateau, Antarctic Peninsula: evidence for recent climate change. *Ann. Glaciol.* **20,** 420–426 (1994).
21. Yuan, X. J. ENSO-related impacts on Antarctic sea ice: a synthesis of phenomenon and mechanisms. *Antarct. Sci.* **16,** 415–425 (2004).
22. Mann, M. E. *et al.* Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proc. Natl Acad. Sci. USA* **105,** 13252–13257 (2008).
23. Bracegirdle, T. J., Connolley, W. M. & Turner, J. Antarctic climate change over the twenty first century. *J. Geophys. Res.* **113,** D03103 (2008).
24. Hodgson, D. A. First synchronous retreat of ice shelves marks a new phase of polar deglaciation. *Proc. Natl Acad. Sci. USA* **108,** 18859–18860 (2011).
25. Aristarain, A. J., Delmas, R. J. & Stievenard, M. Ice-core study of the link between sea-salt aerosol, sea-ice cover and climate in the Antarctic Peninsula area. *Clim. Change* **67,** 63–86 (2004).
26. Sime, L. C., Tindall, J. C., Wolff, E. W., Connolley, W. M. & Valdes, P. J. Antarctic isotopic thermometer during a $CO_2$ forced warming event. *J. Geophys. Res.* **113,** D24119 (2008).
27. Hansen, J., Ruedy, R., Sato, M. & Lo, K. Global surface temperature change. *Rev. Geophys.* **48,** RG4004 (2010).
28. Smith, T. M., Reynolds, R. W., Peterson, T. C. & Lawrimore, J. Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006). *J. Clim.* **21,** 2283–2296 (2008).
29. EPICA Community Members. Eight glacial cycles from an Antarctic ice core. *Nature* **429,** 623–628 (2004).
30. EPICA Community Members. One-to-one coupling of glacial climate variability in Greenland and Antarctica. *Nature* **444,** 195–198 (2006).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank our colleague in the field, S. Shelley, who took part in the ice-core drilling project; the captain and crew of HMS *Endurance*, who provided logistical support for the drilling field season; S. Kipfstuhl and the Alfred Wegner Institute at Bremerhaven for assistance in the processing of the ice core; J. Smellie and S. Roberts for discussions on Antarctic Peninsula tephras; D. Hodgson and E. Wolff for comments during preparation of the manuscript; and E. Capron, N. Lang, J. Levine and E. Ludlow for laboratory assistance. This study is part of the British Antarctic Survey Polar Science for Planet Earth Programme and was funded by the Natural Environment Research Council. Support from the Institut Polaire Français - Paul Emile Victor (IPEV), and from the Institut National des Sciences de l'Univers in France (INSU/PNEDC "AMANCAY" project), facilitated by J. Chappellaz and F. Vimeux, enabled the technical contribution of the French National Center for Drilling and Coring (INSU/C2FN).

**Author Contributions** R.M. designed the project. R.M., N.J.A. and R.C.A.H. constructed the age scale, and R.M., N.J.A., C.A., L.F. and J.T. performed the isotopic, chemical and physical measurements to characterize the ice. R.M., N.J.A., J.T., L.C.S., O.A. and S.F. were involved with the logistics and fieldwork that enabled the ice-core drilling. R.M. and N.J.A. co-wrote the manuscript.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.M. (rmu@bas.ac.uk).

doi:10.1038/nature11345

# Dopamine neurons modulate pheromone responses in *Drosophila* courtship learning

Krystyna Keleman[1], Eleftheria Vrontou[1]†, Sebastian Krüttner[1], Jai Y. Yu[1]†, Amina Kurtovic-Kozaric[1]† & Barry J. Dickson[1]

**Learning through trial-and-error interactions allows animals to adapt innate behavioural 'rules of thumb' to the local environment, improving their prospects for survival and reproduction. Naive *Drosophila melanogaster* males, for example, court both virgin and mated females, but learn through experience to selectively suppress futile courtship towards females that have already mated[1]. Here we show that courtship learning reflects an enhanced response to the male pheromone *cis*-vaccenyl acetate (cVA), which is deposited on females during mating and thus distinguishes mated females from virgins. Dissociation experiments suggest a simple learning rule in which unsuccessful courtship enhances sensitivity to cVA. The learning experience can be mimicked by artificial activation of dopaminergic neurons, and we identify a specific class of dopaminergic neuron that is critical for courtship learning. These neurons provide input to the mushroom body (MB) γ lobe, and the DopR1 dopamine receptor is required in MBγ neurons for both natural and artificial courtship learning. Our work thus reveals critical behavioural, cellular and molecular components of the learning rule by which *Drosophila* adjusts its innate mating strategy according to experience.**

Mature virgin *Drosophila* females are usually willing to mate, whereas those that have recently mated are generally recalcitrant to further mating attempts. A male thus increases his overall mating success if he concentrates his courtship efforts on virgins. Given geographic and seasonal fluctuations in the relative abundance of virgins and mated females, and the cues that distinguish them, the optimal courtship strategy is unlikely to be a species universal. A heuristic for approaching this optimum could, however, be universal, allowing evolution to select for genes that implement such a learning rule in the fly's brain.

A male's courtship behaviour can be quantified by a courtship index (CI), and his ability to discriminate virgins from mated females by a discrimination index (DI), the relative reduction in the mean CI in single-pair assays with mated versus virgin females: $\mathrm{DI} = [\mathrm{CI}_v - \mathrm{CI}_m]/\mathrm{CI}_v$. In our assays, naive males courted mated females only marginally less vigorously than they courted virgins (DI = 13.8%; Fig. 1a, b and Supplementary Table 1a), whereas males that had experienced rejection from mated females were subsequently much less active when courting mated females than virgins (DI = 51.6%; Fig. 1a, b and Supplementary Table 1a). The relative difference between the mean CIs of experienced ($\mathrm{CI}^+$) and naive ($\mathrm{CI}^-$) males gives rise to a learning index: $\mathrm{LI} = [\mathrm{CI}^- - \mathrm{CI}^+]/\mathrm{CI}^-$. For males trained with mated females, the LI was just 7.8% in tests with virgin females but 48.2% when tested with mated females (Fig. 1c, d and Supplementary Table 1b). Similar results were obtained when decapitated virgins were used as trainers (Fig. 1e, f and Supplementary Table 2), suggesting that male behaviour is conditioned by the failure to mate, not by active rejection from the female.

To discriminate mated females from virgins, a male might detect either the subtle changes in female pheromones on mating[2] or the telltale vestiges of male pheromones that linger on mated females[3]. The male-specific pheromone cVA is transferred to the female cuticle on mating[3–5]. It is not detectable on the cuticle of either males or virgin females[5]. Naive *Or67d* mutant males, which are unable to detect cVA[6–8], courted virgin and mated females equally (DI = −0.4%) and did not benefit from training (LI = −3.0%; Fig. 1g, h and Supplementary Tables 3 and 4). In contrast, analogous mutations in either of two other candidate pheromone receptor genes[9,10], *Or47b* and *Gr68a*, did not impair discrimination or learning (Fig. 1g, h and Supplementary Tables 3 and 4). cVA detection is therefore crucial for naive and experienced males to discriminate mated females from virgins.

The salient feature of training might be the presence of cVA on the mated female, the lack of courtship success, or an association formed between the two. We designed a dissociation experiment to distinguish between these possibilities. Female post-mating behaviour, including courtship rejection, is triggered by sex peptide (SP), a male seminal fluid peptide transferred to the female during mating[11]. Virgin females in which SP is transgenically expressed in the nervous system reject courting males[12] (pseudomated females), whereas females that have mated with *SP*-null mutant males are still receptive[13] (pseudovirgins). As expected, we detected cVA on the cuticle of both mated females and pseudovirgins (178.8 ± 11.0 and 57.5 ± 14.7 ng per fly (means ± s.e.m.), respectively; $n = 3$), but not on virgins or pseudomated females ($n = 3$). Thus, with pseudomated and pseudovirgin females the presence of cVA and sexual receptivity are fully dissociated.

Pseudomated females were just as effective as genuinely mated females when used as trainers (Fig. 1i, j and Supplementary Table 5), whereas pseudovirgin females were not (Fig. 1k, l and Supplementary Table 6). In contrast, pseudovirgin but not pseudomated females were as effective as mated females when used as testers (Fig. 1i–l and Supplementary Tables 5 and 6). Indeed, robust courtship learning was observed when males were trained with pseudomated females and tested with pseudovirgins, but not vice versa (Fig. 1m, n and Supplementary Table 7). We therefore conclude that the salient feature of training is simply the lack of courtship success, not its association with cVA, and that training alters the male's response to cVA or some other vestige of previous contact with another male.

To test whether training does indeed alter sensitivity to cVA, we applied varying doses of cVA to pseudomated females and presented them as testers to naive and experienced males. As expected[3,6], high doses of cVA inhibited courtship by both naive and experienced males (Fig. 1o and Supplementary Table 8). However, males trained with either mated or pseudomated females were inhibited by much lower doses of cVA than naive males were (Fig. 1o and Supplementary Table 8). Courtship training did not enhance sensitivity to an unrelated aversive odorant (Supplementary Fig. 1).

Dopamine is thought to provide a learning signal in a variety of different models and species, including aversive olfactory learning[14,15] and conditioned suppression of male–male courtship[16] in *Drosophila*. If dopamine also encodes an instructive signal during courtship learning,

---

[1] Research Institute of Molecular Pathology, Dr Bohrgasse 7, A-1030 Vienna, Austria. †Present addresses: Department of Physiology, Anatomy and Genetics, University of Oxford, Sherrington Building, Parks Road, Oxford OX1 3PT, UK (E.V.); University of California at San Francisco (UCSF) Center for Integrative Neurosciences, 675 Nelson Rising Lane, San Francisco, California 94143-0444, USA (J.Y.Y.); Department of Pathology, Clinical Center of the University of Sarajevo, Bolnicka 35, 71000 Sarajevo, Bosnia and Herzegovina (A.K.-K.).

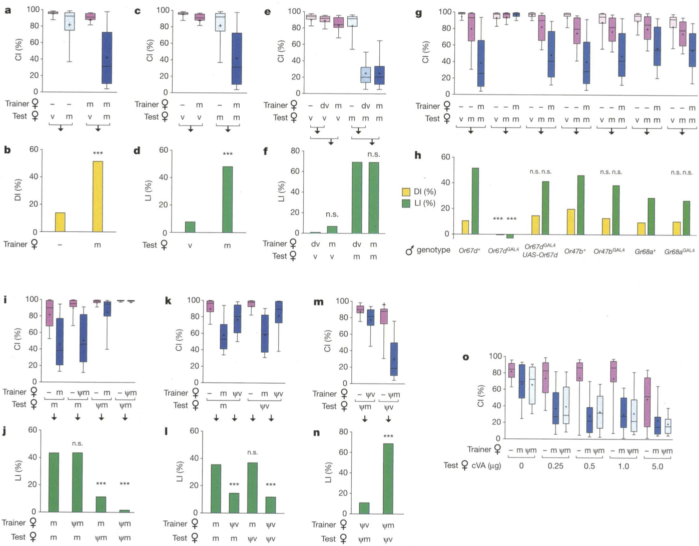


**Figure 1 | Experience enhances the behavioural response to cVA.** **a–f**, Courtship (**a**, **c**, **e**), discrimination (**b**) and learning indices (**d**, **f**) of wild-type males. Trainer and tester females: –, none (naive males); m, mated female; v, virgin; dv, decapitated virgin. Box-and-whisker plots for CI show 10th, 25th, 50th, 75th and 90th centiles and mean (+). Three asterisks, $P < 0.001$ compared with naive male (**b**) or virgin tester (**d**); n.s., $P > 0.05$ compared with decapitated virgin trainers (**f**). **g**, **h**, Courtship index (**g**) and discrimination and learning indices (**h**) of *Or67d*, *Or47b* and *Gr68a* mutant males. n.s., $P > 0.05$; three asterisks, $P < 0.001$ compared with wild-type controls. **i–m**, Courtship (**i**, **k**, **m**) and learning (**j**, **l**, **n**) indices of wild-type males in dissociation experiments using pseudomated females (ψm, *elav-GAL4 UAS-SP*) and pseudovirgin females (ψv, wild-type females previously mated to *SP*-null mutant males). n.s., $P > 0.05$; three asterisks, $P < 0.001$ compared with assays with mated females as trainers and testers (**j**, **l**), or the reciprocal assay (**n**). Post-mating behaviours are not completely eliminated in pseudovirgin females, because SP function can be partly compensated for by the related DUP99B peptide[29]. **o**, Courtship indices of naive and experienced males towards pseudomated females perfumed with varying doses of cVA. $P < 0.01$ for all comparisons of experienced to naive males; $P > 0.05$ for all comparisons between males trained with mated versus pseudomated females.

then artificial stimulation of dopaminergic neurons might mimic training with a mated female. To test this, we expressed the warmth-activated TrpA1 channel[17] in most dopaminergic neurons[18], and attempted to 'train' naive isolated males by warming them briefly to 30 °C. When subsequently returned to 25 °C and tested with mated females, the courtship activity of these males was indeed markedly reduced in comparison with that of control males (Fig. 2a, b and Supplementary Table 9). This suppression was specific for courtship towards mated but not virgin females, was dependent on a functional Or67d receptor (Fig. 2a, b and Supplementary Table 9), and was correlated with an increased sensitivity to cVA (Fig. 2c and Supplementary Table 10). In these respects, activation of dopaminergic neurons thus mimics a specific courtship learning signal rather than a non-specific punishment signal that might be expected to suppress courtship more generally. Experiments in which we selectively activated various subsets of dopaminergic neurons further suggest that the neurons involved in courtship learning are distinct from those previously implicated in various forms of aversive olfactory learning[19,20] (Fig. 2d, e and Supplementary Table 11).

Many aspects of male courtship behaviour have been linked to the set of neurons that express the *fruitless* (*fru*) gene[21]. Among these are the Or67d olfactory neurons (OSNs) and MB$\gamma$ neurons, both of which function in courtship learning (Fig. 1e, f and ref. 22). We speculated that the dopaminergic neurons involved in courtship learning might also be *fru*$^+$. To test this hypothesis we acutely blocked synaptic transmission of *fru*$^+$ dopaminergic neurons by using *shi*$^{ts}$ (refs 23–25), which inhibits synaptic vesicle recycling at 30 °C but not at 22 °C. Such males showed significantly impaired learning when trained at 30 °C and tested at 22 °C, but not vice versa (Fig. 2f, g and Supplementary Table 12). These data thus establish a requirement for dopaminergic neurons in memory formation, not recall, and further indicate that the relevant cells are *fru*$^+$.

We previously identified two distinct classes of *fru*$^+$ dopaminergic neurons: aSP4 and aSP13 (ref. 25) (Fig. 3a–g and Supplementary Fig. 2).

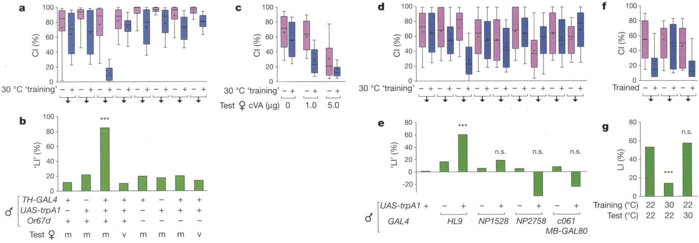


**Figure 2 | Activation of dopaminergic neurons is necessary and sufficient for learning. a, b,** Courtship (**a**) and 'fictive learning' (**b**) indices of males of the indicated genotypes. Before testing, isolated males were either retained at the normal culture temperature of 22 °C (−) or warmed to 30 °C for 45 min (+). Three asterisks, $P < 0.001$ compared with *TH-GAL4*/+ *Or67d*$^+$ males. **c**, Courtship indices of naive and fictively trained males towards pseudomated females perfumed with various doses of cVA. $P < 0.01$ for all comparisons at a given cVA dose. **d, e**, Courtship (**d**) and 'fictive learning' (**e**) indices of males of the indicated genotypes. Three asterisks, $P < 0.001$; n.s., $P > 0.05$ compared with corresponding control without *UAS-trpA1*. *HL9-GAL4* includes most dopaminergic neurons, but not PPL1 cluster neurons implicated in olfactory learning in a heat punishment assay[19]. The other *GAL4* lines drive expression in MB-M3 or MB-MP1 neurons, implicated in olfactory learning in an electric shock model[20]. **f, g**, Courtship (**f**) and learning (**g**) indices of *fru*$^{\mathrm{FLP}}$ *TH-GAL4 UAS>stop>shi*$^{\mathrm{ts}}$ males. Training and testing were performed at the indicated temperatures with mated females. Box-and-whisker plots for CI show 10th, 25th, 50th, 75th and 90th centiles and mean (+). Two asterisks, $P < 0.01$; n.s., $P > 0.05$ compared with males trained and tested at 22 °C.

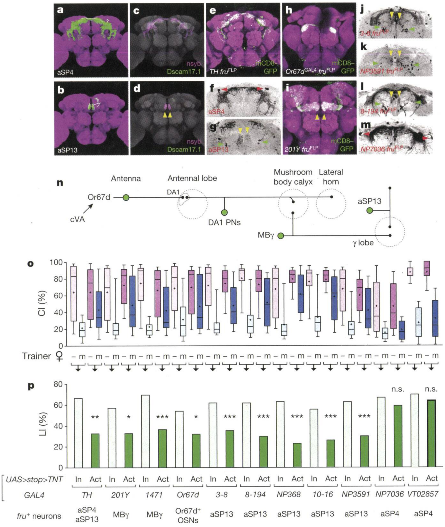

a
aSP4
b
aSP13
c
nsyb
Dscam17.1
d
e
TH fru^FLP
mCD8-GFP
f
aSP4
g
aSP13
h
Or67d^GAL4 fru^FLP
mCD8-GFP
i
201Y fru^FLP
mCD8-GFP
j
3-8 fru^FLP
k
NP3591 fru^FLP
l
8-194 fru^FLP
m
NP7036 fru^FLP
n
Antenna
Antennal lobe
Mushroom body calyx
Lateral horn
Or67d
cVA
DA1
DA1 PNs
aSP13
MBγ
γ lobe
o
CI (%)
Trainer ♀
p
LI (%)
** * *** * *** *** *** *** *** n.s. n.s.
UAS>stop>TNT
In Act
GAL4
TH 201Y 1471 Or67d 3-8 8-194 NP368 10-16 NP3591 NP7036 VT02857
fru+ neurons
aSP4 aSP13
MBγ
MBγ
Or67d+ OSNs
aSP13
aSP4

**Figure 3 | Courtship learning requires synaptic transmission of aSP13 neurons. a, b**, Surface representation of aSP4 (**a**) and aSP13 (**b**) neurons in a male brain[25]. There is typically one aSP4 neuron and two to four aSP13 neurons per hemisphere. **c, d**, Overlay of registered and masked confocal images[25] of aSP4 (**c**) and aSP13 (**d**) neurons, labelled with the presynaptic marker green fluorescent protein (GFP)-tagged nsyb (magenta) and the dendritic marker Dscam17.1–GFP (green). Yellow arrowheads in **d** indicate the presynaptic innervation of aSP13 at the tip of the MB γ lobe. **e**, Brain of a *TH-GAL4 fru*$^{\mathrm{FLP}}$ *UAS>stop>mCD8-GFP* male stained with anti-GFP (green) and the general synaptic marker monoclonal antibody nc82 (magenta). **f, g**, Enlarged and inverted views of the green channel of **e**. Arrowheads indicate aSP4 (red, **f**) and aSP13 (green, **g**) soma. **h, i**, Brain of *fru*$^{\mathrm{FLP}}$ *UAS>stop>mCD8-GFP* males carrying either *Or67d*$^{\mathrm{GAL4}}$ (**h**) or *201Y-GAL4* (**i**), stained with anti-GFP (green) and monoclonal antibody nc82 (magenta). **j–m**, Brains of *fru*$^{\mathrm{FLP}}$ *UAS>stop>mCD8-GFP* males, additionally carrying the indicated *GAL4* driver, stained with anti-GFP (black). **n**, Diagram of cVA processing pathway, adapted from ref. 25. PN, olfactory projection neuron. **o, p**, Courtship (**o**) and learning (**p**) indices of males carrying *fru*$^{\mathrm{FLP}}$, the indicated GAL4 driver and either the active (Act) or inactive (In) versions of *UAS>stop>TNT*. Box-and-whisker plots for CI show 10th, 25th, 50th, 75th and 90th centiles and mean (+). n.s., $P > 0.05$; asterisk, $P < 0.05$; two asterisks, $P < 0.01$; three asterisks, $P < 0.001$ compared with corresponding control with inactive *TNT* transgene.

To test whether aSP4 and/or aSP13 neurons contribute to courtship learning, we chronically inhibited synaptic transmission in these neurons with tetanus toxin light chain (TNT), using drivers selective for either aSP4 or aSP13 (refs 24, 25). With each of five independent aSP13 drivers, learning was reduced by about 50% compared with control males that carried an inactive version of the *TNT* transgene in the same genetic background (Fig. 3j–l, o, p and Supplementary Table 13). A similar learning deficit was observed in positive controls in which we targeted TNT to both aSP13 and aSP4, to *Or67d*$^+$ OSNs[6], or to MBγ neurons[26,27] (Fig. 3h, i, o, p and Supplementary Table 13). In contrast, courtship learning was unimpaired in assays using either of two driver lines expressed in aSP4 but not aSP13 (Fig. 3m, o, p and Supplementary Table 13). We conclude that synaptic transmission of aSP13 neurons is crucial for courtship learning.

The presynaptic termini of aSP13 neurons are located at the tip of the MB γ lobe (Fig. 3d), indicating that they might convey a dopamine learning signal to MBγ neurons. If so, then a dopamine receptor should be required specifically in MBγ neurons for courtship learning. We considered the DopR1 and DopR2 receptors as candidates, and used homologous recombination to generate analogous loss-of-function alleles for each gene (*DopR1*$^{\mathrm{attP}}$ and *DopR2*$^{\mathrm{attP}}$, respectively). Both mutants are viable and fertile and homozygous naive males court at normal levels (Fig. 4a and Supplementary Table 14). However, courtship learning was significantly impaired in *DopR1*$^{\mathrm{attP}}$ but not *DopR2*$^{\mathrm{attP}}$ mutants (Fig. 4b and Supplementary Table 14), as was 'fictive learning' induced by thermogenetic activation of dopaminergic neurons (Fig. 4c, d and Supplementary Table 15). Nevertheless, learning was not completely eliminated in these *DopR1* mutants, indicating that other dopamine receptors might also contribute. To confirm that the learning deficit in the *DopR1*$^{\mathrm{attP}}$ mutant was indeed due to loss of *DopR1* function, we reintegrated the deleted genomic region by site-specific transgenesis. Males homozygous for this repaired *DopR1*

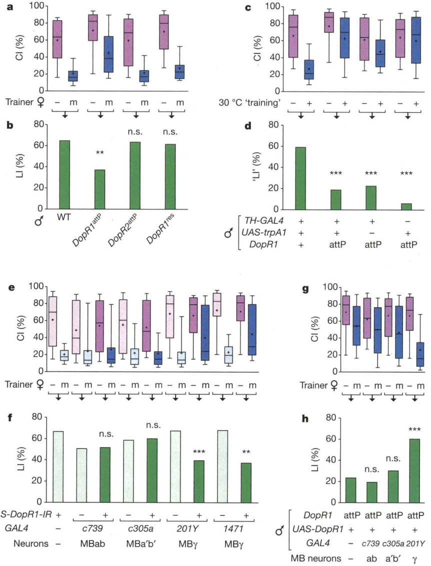


**Figure 4 | DopR1 functions in MBγ neurons. a, b,** Courtship (**a**) and learning (**b**) indices of DoR mutants. n.s., $P > 0.05$; two asterisks, $P < 0.01$ compared with wild-type (WT) males. **c, d,** Courtship (**c**) and 'fictive learning' (**d**) indices in fictive learning assays with mated female testers. Before testing, isolated males were either retained at the normal culture temperature of 22 °C (–) or warmed to 30 °C for 45 min (+). For male genotypes, + and – indicate the presence or absence, respectively, of the *TH-GAL4* and *UAS-trpA1* transgenes; for *DopR1*, '+' indicates the wild-type control allele and 'attP' the *DopR1*$^{\mathrm{attP}}$ mutant. Three asterisks, $P < 0.0001$ compared to wild type males. **e, f,** Courtship (**e**) and learning (**f**) indices on RNAi knockdown of *DopR1*. For male genotypes, + and – indicate the presence or absence of the *UAS-DopR1-IR* RNAi transgene. n.s., $P > 0.05$; two asterisks, $P < 0.01$; three asterisks, $P < 0.0001$ compared with control males without the *UAS-DopR1-IR* transgene. **g, h,** Courtship (**g**) and learning (**h**) indices on rescue of *DopR1* function. All males are *DopR1*$^{\mathrm{attP}}$ mutants (attP) carrying a *UAS-DopR1* transgene (+) and either no (–) or the indicated *GAL4* driver. Box-and-whisker plots for CI show 10th, 25th, 50th, 75th and 90th centiles and mean (+). n.s., $P > 0.05$; three asterisks, $P < 0.0001$ compared with control males without a *GAL4* driver.

allele, *DopR1*$^{\mathrm{Res}}$, performed just as well as wild-type males in courtship learning assays (Fig. 4a, b and Supplementary Table 14).

Finally, we performed RNA-mediated interference (RNAi) knockdown and rescue experiments to test whether DopR1 function is indeed required in MB$\gamma$ neurons. Expression of a *DopR1* RNAi transgene selectively in MB$\gamma$ neurons significantly reduced DopR1 expression levels in the $\gamma$ lobe (Supplementary Fig. 3) and impaired courtship learning (Fig. 4e, f and Supplementary Table 16). Conversely, the learning disability of *DopR1*$^{\mathrm{attP}}$ mutants was fully alleviated by expressing a *DopR1* transgene specifically in MB$\gamma$ neurons (Fig. 4g, h and Supplementary Table 17). We therefore postulate that DopR1 acts in MB$\gamma$ neurons to transduce a dopamine learning signal provided by aSP13 neurons.

To maximize his reproductive success, a *Drosophila* male should be highly attuned to those cues that discriminate receptive from unreceptive females. A male that is too selective may miss mating opportunities; a male that is too promiscuous may waste resources on futile courtship. The optimal tuning is likely to vary from place to place and from time to time, depending for example on local and seasonal fluctuations in the abundance and quality of mating partners and the pheromone signals that they provide. Our study defines a simple heuristic that could allow the male to learn an effective courtship strategy in his local environment: be promiscuous at first, but become more selective if a mating attempt fails. Furthermore, we have identified key elements that implement this learning rule in the fly's brain. We propose that, when a mating attempt fails, aSP13 dopaminergic neurons convey a learning signal to MB$\gamma$ neurons through the DopR1 receptor, and that this induces lasting changes in the internal processing of the cVA signal that discriminates mated females from virgins. Further studies of this genetically defined and tractable circuit should provide a detailed understanding of how a relatively simple learning circuit, embedded within decision-making centres of the brain, endows plasticity on an innate behaviour.

## METHODS SUMMARY

Courtship conditioning assays and data analyses were performed as described previously[22]. CIs, defined as the percentage of time for which the male courts the female during a 10-min observation period, were scored manually from video recordings. Mann–Whitney–Wilcoxon tests were used for statistical comparisons of CIs between two data sets. Permutations tests were used to compare DIs and LIs, with 100,000 permutations of the raw data. For 'fictive training', males were collected at eclosion and aged in isolation for 5–7 days at 22 °C, transferred by gentle aspiration to prewarmed chambers at 30 °C for 45 min, then to 25 °C for 10–25 min before testing. Perfuming experiments with cVA were performed by applying 1 µl of appropriate dilution to the female's abdomen about 45 min before use as a tester. Immunostaining, confocal microscopy, image registration, and visualization were performed as described[25]. *Or47b*$^{\mathrm{GAL4}}$ and *Gr68a*$^{\mathrm{GAL4}}$ alleles were generated by ends-in homologous recombination, and *DopR1*$^{\mathrm{attP}}$ and *DopR2*$^{\mathrm{attP}}$ by ends-out targeting[28]. For the *Or47b* and *Gr68a* mutants the GAL4-coding region replaces the entire endogenous coding region. For the *DopR1* and *DopR2* mutants the *attP* site replaces the respective first coding exons. The *UAS-DopR1-IR* line is from the KK library maintained at the Vienna *Drosophila* RNAi Center (VDRC; http://www.vdrc.at).

**Full Methods** and any associated references are available in the online version of the paper.

**Received 18 January 2011; accepted 25 June 2012.**

**Published online 19 August 2012.**

1. Siegel, R. W. & Hall, J. C. Conditioned responses in courtship behavior of normal and mutant *Drosophila*. *Proc. Natl Acad. Sci. USA* **76,** 3430–3434 (1979).
2. Tompkins, L. Genetic analysis of sex appeal in *Drosophila*. *Behav. Genet.* **14,** 411–440 (1984).
3. Jallon, J. M., Antony, C. & Benamar, O. Un anti-aphrodisiaque produit part les males de *Drosophila* et transféré aux femelles lors de la copulation. *C. R. Acad. Sci. Paris* **292,** 1147–1149 (1981).
4. Butterworth, F. M. Lipids of *Drosophila*: a newly detected lipid in the male. *Science* **163,** 1356–1357 (1969).
5. Everaerts, C., Farine, J. P., Cobb, M. & Ferveur, J. F. *Drosophila* cuticular hydrocarbons revisited: mating status alters cuticular profiles. *PLoS ONE* **5,** e9607 (2010).
6. Kurtovic, A., Widmer, A. & Dickson, B. J. A single class of olfactory neurons mediates behavioural responses to a *Drosophila* sex pheromone. *Nature* **446,** 542–546 (2007).
7. Ha, T. S. & Smith, D. P. A pheromone receptor mediates 11-*cis*-vaccenyl acetate-induced responses in *Drosophila*. *J. Neurosci.* **26,** 8727–8733 (2006).
8. van der Goes van Naters, W. & Carlson, J. R. Receptors and neurons for fly odors in *Drosophila*. *Curr. Biol.* **17,** 606–612 (2007).
9. Root, C. M., Semmelhack, J. L., Wong, A. M., Flores, J. & Wang, J. W. Propagation of olfactory information in *Drosophila*. *Proc. Natl Acad. Sci. USA* **104,** 11826–11831 (2007).
10. Bray, S. & Amrein, H. A putative *Drosophila* pheromone receptor expressed in male-specific taste neurons is required for efficient courtship. *Neuron* **39,** 1019–1029 (2003).
11. Chen, P. S. *et al.* A male accessory gland peptide that regulates reproductive behavior of female *D. melanogaster*. *Cell* **54,** 291–298 (1988).
12. Nakayama, S., Kaiser, K. & Aigaki, T. Ectopic expression of sex-peptide in a variety of tissues in *Drosophila* females using the *P[GAL4]* enhancer-trap system. *Mol. Gen. Genet.* **254,** 449–455 (1997).
13. Liu, H. & Kubli, E. Sex-peptide is the molecular basis of the sperm effect in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **100,** 9929–9933 (2003).
14. Wise, R. A. Dopamine, learning and motivation. *Nature Rev. Neurosci.* **5,** 483–494 (2004).
15. Waddell, S. Dopamine reveals neural circuit mechanisms of fly memory. *Trends Neurosci.* **33,** 457–464 (2010).
16. Neckameyer, W. S. Dopamine and mushroom bodies in *Drosophila*: experience-dependent and -independent aspects of sexual behavior. *Learn. Mem.* **5,** 157–165 (1998).
17. Hamada, F. N. *et al.* An internal thermal sensor controlling temperature preference in *Drosophila*. *Nature* **454,** 217–220 (2008).
18. Friggi-Grelin, F. *et al.* Targeted gene expression in *Drosophila* dopaminergic cells using regulatory sequences from tyrosine hydroxylase. *J. Neurobiol.* **54,** 618–627 (2003).
19. Claridge-Chang, A. *et al.* Writing memories with light-addressable reinforcement circuitry. *Cell* **139,** 405–415 (2009).
20. Aso, Y. *et al.* Specific dopaminergic neurons for the formation of labile aversive memory. *Curr. Biol.* **20,** 1445–1451 (2010).
21. Dickson, B. J. Wired for sex: the neurobiology of *Drosophila* mating decisions. *Science* **322,** 904–909 (2008).
22. Keleman, K., Kruttner, S., Alenius, M. & Dickson, B. J. Function of the *Drosophila* CPEB protein Orb2 in long-term courtship memory. *Nature Neurosci.* **10,** 1587–1593 (2007).
23. Kitamoto, T. Conditional modification of behavior in *Drosophila* by targeted expression of a temperature-sensitive *shibire* allele in defined neurons. *J. Neurobiol.* **47,** 81–92 (2001).
24. Stockinger, P., Kvitsiani, D., Rotkopf, S., Tirian, L. & Dickson, B. J. Neural circuitry that governs *Drosophila* male courtship behavior. *Cell* **121,** 795–807 (2005).
25. Yu, J. Y., Kanai, M. I., Demir, E., Jefferis, G. S. & Dickson, B. J. Cellular organization of the neural circuit that drives *Drosophila* courtship behavior. *Curr. Biol.* **20,** 1602–1614 (2010).
26. Zars, T., Fischer, M., Schulz, R. & Heisenberg, M. Localization of a short-term memory in *Drosophila*. *Science* **288,** 672–675 (2000).
27. Isabel, G., Pascual, A. & Preat, T. Exclusive consolidated memory phases in *Drosophila*. *Science* **304,** 1024–1027 (2004).
28. Rong, Y. S. & Golic, K. G. Gene targeting by homologous recombination in *Drosophila*. *Science* **288,** 2013–2018 (2000).
29. Saudan, P. *et al.* Ductus ejaculatorius peptide 99B (DUP99B), a novel *Drosophila melanogaster* sex-peptide pheromone. *Eur. J. Biochem.* **269,** 989–997 (2002).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** We thank É. Leitner for performing the gas chromatography–mass spectrometry analysis; P. Garrity, U. Heberlein, M. Heisenberg, E. Kubli, S. Waddell, the *Drosophila* Genetic Resource Centre, the Bloomington Stock Center and the VDRC for fly stocks; K. Jandrasits and Z. Portik-Dobos for technical assistance; M. Zimmer for critical comments on the manuscript. Basic research at the Institute of Molecular Pathology is funded in part by Boehringer Ingelheim GmbH. This work was additionally supported by grants from the European Research Council (B.J.D.) and the Austrian Science Fund (K.K.). E.V. was supported by a European Molecular Biology Organization long-term postdoctoral fellowship.

**Author Contributions** K.K. and B.J.D. designed the experiments and performed the data analysis. K.K. performed most behavioural experiments. B.J.D. wrote the manuscript together with K.K. E.V. generated the *DopR1* and *DopR2* mutants, and A.K. generated the *Or47b* and *Gr68a* mutants. S.K. and J.Y.Y. performed the antibody stainings.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.K. (keleman@imp.ac.at) or B.J.D. (dickson@imp.ac.at).

LETTER

doi:10.1038/nature11306

# Neuronal circuitry mechanism regulating adult quiescent neural stem-cell fate decision

Juan Song[1,2], Chun Zhong[1,2], Michael A. Bonaguidi[1,2], Gerald J. Sun[1,3], Derek Hsu[1], Yan Gu[4], Konstantinos Meletis[5], Z. Josh Huang[6], Shaoyu Ge[4], Grigori Enikolopov[6], Karl Deisseroth[7], Bernhard Luscher[8], Kimberly M. Christian[1,2], Guo-li Ming[1,2,3] & Hongjun Song[1,2,3]

**Adult neurogenesis arises from neural stem cells within specialized niches[1–3]. Neuronal activity and experience, presumably acting on this local niche, regulate multiple stages of adult neurogenesis, from neural progenitor proliferation to new neuron maturation, synaptic integration and survival[1,3]. It is unknown whether local neuronal circuitry has a direct impact on adult neural stem cells. Here we show that, in the adult mouse hippocampus, nestin-expressing radial glia-like quiescent neural stem cells[4–9] (RGLs) respond tonically to the neurotransmitter $\gamma$-aminobutyric acid (GABA) by means of $\gamma_2$-subunit-containing GABA$_A$ receptors. Clonal analysis[9] of individual RGLs revealed a rapid exit from quiescence and enhanced symmetrical self-renewal after conditional deletion of $\gamma_2$. RGLs are in close proximity to terminals expressing 67-kDa glutamic acid decarboxylase (GAD67) of parvalbumin-expressing (PV$^+$) interneurons and respond tonically to GABA released from these neurons. Functionally, optogenetic control of the activity of dentate PV$^+$ interneurons, but not that of somatostatin-expressing or vasoactive intestinal polypeptide (VIP)-expressing interneurons, can dictate the RGL choice between quiescence and activation. Furthermore, PV$^+$ interneuron activation restores RGL quiescence after social isolation, an experience that induces RGL activation and symmetrical division[8]. Our study identifies a niche cell-signal-receptor trio and a local circuitry mechanism that control the activation and self-renewal mode of quiescent adult neural stem cells in response to neuronal activity and experience.**

Recent genetic lineage-tracing studies have identified nestin-expressing RGLs as quiescent neural stem cells (qNSCs) in the adult mouse hippocampus[4–9]. In adult *nestin-GFP* mice[10], cells expressing green fluorescent protein (GFP$^+$ cells) in the subgranular zone (SGZ) with radial processes expressed GFAP (glial fibrillary acidic protein) but rarely MCM2 (minichromosome maintenance type 2), indicating quiescence (Supplementary Fig. 1a, b). To assess whether local interneurons regulate adult qNSCs directly by means of neurotransmitter release, we examined RGL responses to GABA in slices acutely prepared from adult *nestin-GFP* mice by electrophysiology (see Methods). GFP$^+$ RGLs recorded under whole-cell voltage-clamp showed prominent responses to GABA (200 mM) or the GABA$_A$ receptor (GABA$_A$R) agonist muscimol (200 mM), which were abolished by the GABA$_A$R antagonist bicuculline (BMI; 50 μM; Supplementary Fig. 1c, d). GABA responses were potentiated by diazepam (1 μM), which specifically enhances $\gamma_2$-containing GABA$_A$R responses to GABA[11]. Indeed, GFP$^+$ RGLs showed immunoreactivity to $\gamma_2$ (Supplementary Fig. 1e). $\gamma_2$-containing GABA$_A$Rs are present in non-neuronal cells and can be found both outside and inside synapses in mature neurons[11]. No spontaneous or evoked synaptic currents in response to field stimulation of the dentate granule cell layer were detected in GFP$^+$ RGLs ($n = 25$ cells; Supplementary Fig. 1f, g). Instead, tonic GABA responses were recorded ($n = 18$ cells; Fig. 1 and Supplementary Fig. 1g, h), suggesting GABA spill-over from nearby synapses[11]. To exclude the possibility of synaptic inputs with low release probabilities, we applied hypertonic solution to enhance presynaptic release[12]. Increased GABA tonic responses, but not synaptic currents, were observed (Supplementary Fig. 1h). Inhibition of the GABA reuptake transporter GAT1 with NO-711 (10 μM) also increased tonic responses (Fig. 1), further supporting the tonic nature of GABAergic responses in RGLs.

We next explored pharmacological properties of tonic GABA responses in RGLs[13]. Consistent with the $\gamma_2$ involvement, diazepam (1 μM) significantly increased tonic responses, whereas the benzodiazepine antagonist flumazenil (10 μM) decreased them (Fig. 1). The $\alpha_5$-selective benzodiazepine agonist midazolam (10 μM), or the $\beta_3$-selective positive allosteric modulator etomidate (ETMD; 100 nM), increased tonic GABA responses, whereas the $\alpha_5$-selective inverse agonist L-655708 (50 μM) decreased this response (Fig. 1). Together, these results suggest that $\alpha_5\beta_3\gamma_2$ GABA$_A$Rs are present in adult dentate RGLs to mediate tonic responses to GABA.

To examine the functional role of GABA in regulating adult dentate RGLs *in vivo*, we assessed 5-ethynyl-2′-deoxyuridine (EdU) incorporation and MCM2 expression by RGLs after treatment with diazepam (Supplementary Fig. 2a). We identified RGLs as SGZ cells with nestin$^+$ radial processes (Fig. 2a). Stereological quantification showed that treatment with diazepam led to a 45% decrease in the number of EdU$^+$ RGLs compared with vehicle treatment (Fig. 2b). The number

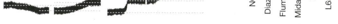


**Figure 1 | Tonic activation of adult quiescent neural stem cells by GABA by means of $\alpha_5\beta_3\gamma_2$ GABA$_A$Rs.** **a**, Sample traces of whole-cell voltage-clamp recording from GFP$^+$ RGLs treated with diazepam (1 μM), flumazenil (10 μM), midazolam (10 μM), ETMD (100 nM) or L-655708 (50 μM), followed by BMI (100 μM) to obtain a baseline for normalizing tonic responses for each cell. **b**, Summary of normalized amplitude of tonic response. Values are means and s.e.m. ($n = 4$ or 5 cells; all significantly different from the basal condition; $P < 0.05$; Student's *t*-test).

---

[1] Institute for Cell Engineering, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [2] Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [3] The Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [4] Department of Neurobiology and Behaviour, State University of New York at Stony Brook, New York 11794, USA. [5] Department of Neuroscience, Karolinska Institutet, S-171 77 Stockholm, Sweden. [6] Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. [7] Department of Bioengineering, Stanford University, Stanford, California 94305, USA. [8] Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA.

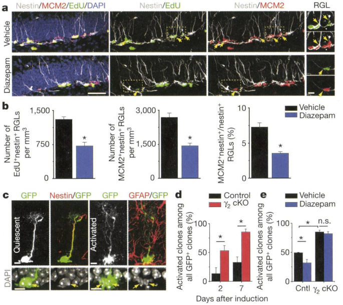

a
Nestin/MCM2/EdU/DAPI
Nestin/EdU
Nestin/MCM2
RGL
Vehicle
Diazepam
b
Number of EdU+nestin+ RGLs per mm3
1,500
750
0
Number of MCM2+nestin+ RGLs per mm3
3,000
1,500
0
MCM2+nestin+/nestin+ RGLs (%)
10
5
0
Vehicle
Diazepam
c
GFP
Nestin/GFP
GFP
GFAP/GFP
Quiescent
Activated
DAPI
d
Control
γ2 cKO
Activated clones among all GFP+ clones (%)
100
50
0
2
7
Days after induction
e
Vehicle
Diazepam
Activated clones among all GFP+ clones (%)
n.s.
Cntl
γ2 cKO

**Figure 2 | Cell-autonomous role of $\gamma_2$-containing $\mathrm{GABA_A}$Rs in maintaining adult neural stem cell quiescence. a, b,** Diazepam promotes quiescence of nestin$^+$ RGLs in the adult dentate gyrus. **a,** Sample confocal images of immunostaining of nestin, MCM2, EdU and 4′,6-diamidino-2-phenylindole (DAPI). Arrows indicate nestin$^+$MCM2$^+$ or nestin$^+$EdU$^+$ RGLs. Scale bars, 50 μm (left) and 10 μm (last column). **b,** Summaries of stereological quantification of RGL EdU incorporation and MCM2 expression. Values are means and s.e.m. ($n = 4$ animals; asterisk, $P < 0.01$; Student's $t$-test). **c–e,** $\gamma_2$ deletion in individual RGLs leads to their activation. **c,** Sample confocal images of immunostaining. Scale bars, 10 μm. **d, e,** Summaries of percentages of RGL clones that were activated (**d**) and those treated with vehicle or diazepam at 7 days after induction (**e**) for control (cntl) and cKO mice. Values are means and s.e.m. ($n = 4$–$8$ animals; asterisk, $P < 0.01$; n.s., $P > 0.1$; Student's $t$-test).

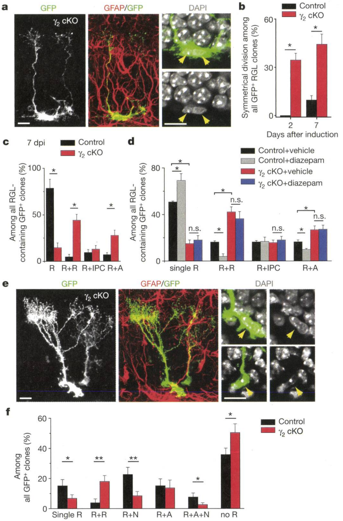

a
GFP
GFAP/GFP
DAPI
γ2 cKO
b
Control
γ2 cKO
Symmetrical division among all GFP+ RGL clones (%)
60
30
0
2
7
Days after induction
c
7 dpi
Control
γ2 cKO
Among all RGL-containing GFP+ clones (%)
100
50
0
R
R+R
R+IPC
R+A
d
Control+vehicle
Control+diazepam
γ2 cKO+vehicle
γ2 cKO+diazepam
Among all RGL-containing GFP+ clones (%)
80
40
0
n.s.
single R
R+R
R+IPC
R+A
e
GFP
GFAP/GFP
DAPI
γ2 cKO
f
Control
γ2 cKO
Among all GFP+ clones (%)
60
40
20
0
*
**
Single R
R+R
R+N
R+A
R+A+N
no R

**Figure 3 | Clonal analysis of RGL fate choice after conditional $\gamma_2$ deletion in individual RGLs in the adult dentate gyrus. a–d,** Short-term effect of $\gamma_2$ deletion on the activation and fate choice of adult dentate RGLs. **a,** Sample confocal images of immunostaining for a GFP$^+$ clone indicating symmetrical division at 7 days after induction. Scale bars, 10 μm. **b–d,** Summaries of percentages of clones indicating symmetrical divisions at 2 and 7 days after induction (**b**), and percentages of different types of RGL clones (**c**) and those treated with vehicle or diazepam (**d**) at 7 days after induction: R + R (two RGLs), R + intermediate progenitor cell (IPC; one RGL and one GFAP$^-$ IPC) and R + A (one RGL and one GFAP$^+$ bushy astrocyte). Values are means and s.e.m. ($n = 4$–$8$ animals; asterisk, $P < 0.05$; n.s., $P > 0.1$; Student's $t$-test). **e, f,** Long-term effect (at 30 days after induction) of $\gamma_2$ deletion on the composition of GFP$^+$ clones in the adult dentate gyrus. **e,** Sample confocal images of immunostaining for a clone consisting of two GFAP$^+$ cells with radial processes. Scale bars, 10 μm. **f,** Summary of percentages of different clone types among all GFP$^+$ clones: R, RGL; N, IPC or neuron; A, astrocyte. Values are means and s.e.m. ($n = 4$–$8$ animals; asterisk, $P < 0.05$; two asterisks, $P < 0.01$; Student's $t$-test).

of MCM2$^+$nestin$^+$ RGLs and the percentage of RGLs that were MCM2$^+$ were also significantly decreased (Fig. 2b). Thus, systemic enhancement of $\gamma_2$-mediated GABA signalling promotes adult dentate RGL quiescence at the population level.

To examine a cell-autonomous role of $\gamma_2$ in RGLs, we generated *nestin-CreER*$^{T2+/-}$;*Z/EG*$^{f/-}$;$\gamma_2^{f/f}$ (cKO) mice and *nestin-CreER*$^{T2+/-}$;*Z/EG*$^{f/-}$;$\gamma_2^{+/+}$ (control) mice and used a low dose of tamoxifen for sparse induction to perform clonal analysis of adult RGLs[9] (Supplementary Fig. 2b–d). Immunohistology and electrophysiology indicated highly efficient, but not complete, $\gamma_2$ deletion in GFP$^+$ RGLs (Supplementary Fig. 2e, f). In cKO mice, the percentage of RGL clones that were activated increased markedly compared with control mice at 2 and 7 days after induction (Fig. 2c, d). Treatment with diazepam decreased the percentage of activated RGL clones in control mice at 7 days after induction, but had no effect in cKO mice (Fig. 2e and Supplementary Fig. 2g). These results showed a direct role of GABA in maintaining adult NSC quiescence through $\gamma_2$ signalling.

We next examined the fate choice of activated RGLs. There was a marked increase in pairs of closely associated GFP$^+$ RGLs at 2 days after induction in adult cKO mice compared with controls, indicating increased RGL symmetrical self-renewal (Fig. 3a, b). Detailed analysis at 7 days after induction showed increased symmetrical and astrogliogenic asymmetrical RGL division in cKO mice (Fig. 3c). Conversely, treatment with diazepam decreased RGL symmetrical division and astrogliogenic asymmetric division in control mice, but had no effect in cKO mice (Fig. 3d). In supporting short-term lineage-tracing results, analysis of clonal composition at 30 days after induction showed decreased percentages of quiescent clones and an increased percentage of clones with multiple RGLs in cKO mice (Fig. 3e, f and Supplementary Fig. 3). Consistent with a role of GABA signalling in promoting new neuron survival[14], percentages of neurogenic clones and multilineage clones were decreased significantly (Fig. 3f and Supplementary Fig. 3e). In contrast, clones without any RGLs were increased in cKO mice (Fig. 3f), suggesting increased RGL depletion after $\gamma_2$ deletion. Together, these gain-of-function and loss-of-function analyses identified GABA as a niche signal to maintain adult NSC quiescence and inhibit symmetrical self-renewal and astrocyte fate choice through $\gamma_2$-containing $\mathrm{GABA_A}$Rs under basal physiological conditions.

We next sought to identify GABA-releasing niche cells among multiple interneuron subtypes in the adult dentate gyrus[15,16]. Immunohistological analysis of adult *nestin–GFP* mice showed a close association between GFP$^+$ RGLs and GAD67$^+$ terminals from PV$^+$

interneurons (Fig. 4a and Supplementary Movie 1). To determine whether PV$^+$ interneurons interact functionally with RGLs, we took an optogenetic approach and used double-floxed (DIO) adeno-associated virus (AAV) to express channelrhodopsin-2 (ChR2) or halorhodopsin (eNpHR3.0) specifically in PV$^+$ interneurons, using adult *PV-Cre* mice[17] (Supplementary Fig. 4a). Immunostaining and electrophysiology confirmed the specificity and efficacy of AAV-mediated opsin expression in controlling the firing of dentate PV$^+$ interneurons (Supplementary Fig. 4b–e). In acute slices from *PV-Cre*$^{+/-}$;*nestin-GFP*$^{+/-}$ mice, photoactivation of PV$^+$ interneurons induced synaptic responses in mature dentate granule cells and tonic responses in GFP$^+$ RGLs to GABA (Fig. 4b, c). Furthermore, a decrease in GABA turnover with the GABA transaminase inhibitor vigabatrin (100 μM) drastically increased GFP$^+$ RGL responses to PV$^+$ interneuron activation (Fig. 4c). Together, these results indicate that adult RGLs respond tonically to GABA released from local PV$^+$ interneurons.

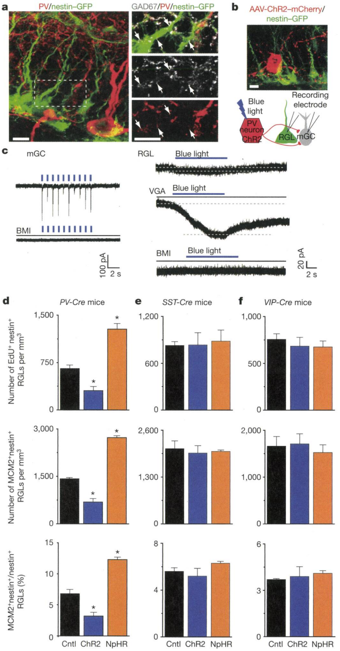


**Figure 4 | Regulation of quiescence and activation state of neural stem cells by PV$^+$, but not SST$^+$ or VIP$^+$ interneuron activity, in the adult dentate gyrus.** **a**, Sample confocal images of GFP and immunostaining of PV and GAD67 (See Supplementary Movie 1). Scale bars, 5 μm. **b**, Sample confocal image and schematic diagram of electrophysiological recording. Scale bar, 10 μm. **c**, Sample whole-cell voltage-clamp recording traces of responses after light stimulation of ChR2$^+$PV$^+$ interneurons from a mature granule cell (mGC; 1 Hz) and a GFP$^+$ RGL (8 Hz) in acute slices, and after treatment with BMI (50 μM) or vigabatrin (VGA; 100 μM). **d–f**, Regulation of RGL activation in the adult dentate gyrus by local interneuron activity. Shown are summaries of stereological quantification of RGL EdU incorporation and MCM2 expression after *in vivo* activation (ChR2) or suppression (NpHR) of PV$^+$ (**d**), SST$^+$ (**e**) or VIP$^+$ (**f**) interneurons or sham treatment (cntl; see Supplementary Figs 5a and 6e for experimental procedures). Values are means and s.e.m. ($n = 3$ or 4 animals; asterisk, $P < 0.01$; Student's $t$-test).

To assess the functional impact of PV$^+$ interneuron activity on RGL behaviour, we photoactivated or suppressed PV$^+$ interneurons in the dentate gyrus of adult *PV-Cre* mice for 5 days (Supplementary Fig. 5a). In comparison with sham treatment without light stimulation, EdU incorporation and MCM2 expression by RGLs were significantly decreased after activation of PV$^+$ interneurons expressing ChR2 tagged with yellow fluorescent protein (ChR2–YFP), resulting in a 53% decrease in RGL activation at the population level (Fig. 4d and Supplementary Fig. 5b). Conversely, suppression of PV$^+$ interneurons expressing eNpHR–YFP led to a 95% increase in RGL activation (Fig. 4d). These results identified PV$^+$ interneurons as a critical niche component and showed that PV$^+$ interneuron activity can dictate the RGL choice between quiescence and activation in the adult dentate gyrus.

Do other subtypes of local interneurons also regulate RGL behaviour *in vivo*? We developed similar optogenetic strategies to manipulate somatostatin-expressing (SST$^+$) or vasoactive intestinal polypeptide-expressing (VIP$^+$) interneurons[16] (Supplementary Fig. 6a). Both SST$^+$ and VIP$^+$ interneurons showed elaborated processes in the SGZ and hilus region (Supplementary Fig. 6c, d and Supplementary Movie 2), and our procedure labelled greater numbers of SST$^+$ and VIP$^+$ interneurons than PV$^+$ interneurons in the adult dentate gyrus (Supplementary Fig. 6b). Electrophysiological recoding of GFP$^+$ RGLs did not detect any tonic or synaptic responses after light-induced activation of SST$^+$ or VIP$^+$ interneurons in acute slices (Supplementary Fig. 6c, d). Functionally, photoactivated or suppressed dentate SST$^+$ or VIP$^+$ interneurons had no effect on EdU incorporation and MCM2 expression by RGLs (Fig. 4e, f and Supplementary Fig. 6e). Thus, coupling of neuronal circuit activity to RGL behaviour seems to be distinctive of PV$^+$ interneurons rather than occurring broadly across different local interneuron subtypes.

Finally, we assessed whether GABA also serves as a niche signal to mediate experience-dependent regulation of RGLs. We subjected mice to a social isolation regime, which decreases neuronal activity in the adult dentate gyrus[18] and was recently shown to promote RGL expansion[8]. Clonal analysis at 7 days after induction showed that, in contrast with group housing, social isolation led to a significant increase in GFP$^+$ RGL activation and symmetrical and astrogenic division, in a similar manner to $\gamma_2$ deletion in RGLs (Fig. 5a, b and Supplementary Fig. 7a). $\gamma_2$-deficient RGLs showed no additional activation or fate alternation after social isolation (Fig. 5b). At the population level, EdU incorporation and MCM2 expression by RGLs were increased significantly after social isolation (Fig. 5c and Supplementary Fig. 7b, c). PV$^+$ interneuron activation abolished the increase in RGL activation induced by social isolation (Fig. 5c). Thus, dentate PV$^+$ interneurons also mediate experience-dependent regulation of adult qNSCs through GABA-$\gamma_2$ signalling.

Precise control of somatic stem cell activity is essential for the long-term maintenance of tissue homeostasis and needs to be closely linked to tissue demands at any given time. Our study of adult RGLs at both clonal and population levels identified a previously unknown niche mechanism that regulates both adult qNSC activation and self-renewal mode in response to neuronal activity and experience (Supplementary Fig. 8). GABA has been shown to decrease the proliferation of other stem cells and progenitors *in vitro*, including mouse embryonic stem

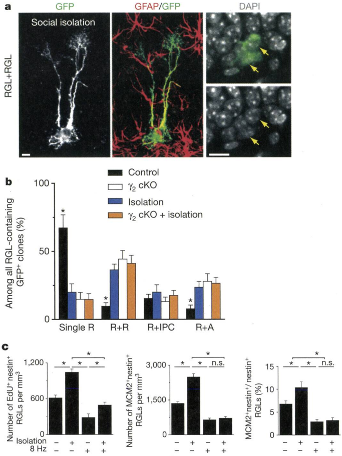


**Figure 5 | Contribution of GABA signalling from PV$^+$ interneurons to experience-dependent regulation of adult quiescent neural stem cells.** **a**, **b**, Clonal analysis of RGL fate choice after social isolation. **a**, Sample confocal images of immunostaining for an activated clone with two RGLs at 7 days after induction after social isolation (see Supplementary Fig. 7 for experimental procedure). Scale bars, 10 μm. **b**, Summary of different types of clone at 7 days after induction. Values are means and s.e.m. ($n = 4$–$8$ animals; asterisk, $P < 0.05$; Student's $t$-test). **c**, Summaries of stereological quantification of RGL EdU incorporation and MCM2 expression. Values are means and s.e.m. ($n = 4$ animals; asterisk, $P < 0.05$; n.s., $P > 0.1$; Student's $t$-test).

cells, by means of GABA$_A$Rs, the phosphatidylinositol-3-OH kinase (PI(3)K)-related kinase family and the histone variant H2AX[19,20]. PTEN deletion in individual RGLs also leads to activation and symmetrical self-renewal in the adult dentate gyrus[9], suggesting a conserved mechanism regulating the proliferation of various stem cells through the GABA$_A$R and PI(3)K/PTEN pathway.

Our optogenetic approach identified PV$^+$ interneurons as a critical and unique niche component among different interneuron subtypes that couples neuronal circuit activity to qNSC regulation *in vivo* under both physiological conditions and in response to specific experience. PV$^+$ interneurons are abundant in the hippocampus and have been implicated in higher brain function and cognitive dysfunction[15]. In the adult dentate gyrus, PV$^+$ interneurons receive excitatory inputs from dentate granule cells and, to a smaller extent, from entorhinal cortical inputs (Supplementary Fig. 8a). We reconstructed one PV$^+$ interneuron in the adult *PV-Cre*$^{+/-}$;*nestin-GFP*$^{+/-}$ mice and estimated that it covered more than 200 GFP$^+$ RGLs in the dentate gyrus (Supplementary Movie 3). A characteristic feature of PV$^+$ interneurons is the formation of ensembles coupled by both electrical (through gap junctions) and chemical connections (through reciprocal innervations)[15]. Thus, PV$^+$ interneurons are well suited to couple local circuit activity to the regulation of a large number of adult NSCs in the hippocampus as an adaptive mechanism—increasing qNSC activation when local circuitry activity levels are low, while keeping NSCs in quiescence when activity levels are high (Supplementary Fig. 8b). Given that both the number and properties of hippocampal PV$^+$ interneurons are regulated by physiological and pathological conditions, such as ageing, Alzheimer's diseases, epilepsy, chronic stress, schizophrenia and other severe psychiatric illness[21–26], our findings have broad implications.

## METHODS SUMMARY

Wild-type (C57BL/6), *nestin-GFP*[10], *PV-Cre*[17], *SST-Cre*[16], *VIP-Cre*[16], *nestin-CreER*$^{T2+/-}$;*Z/EG*$^{f/-}$;$\gamma_2^{f/f}$ (ref. 27) were used in the present study. Cre-dependent recombinant AAV[17] was used for interneuron subtype-specific expression of opsins in the adult dentate gyrus. Electrophysiological recordings and analysis were performed as described previously[28]. Immunohistochemistry, confocal imaging and processing were performed as described previously[9]. Stereological quantification was assessed as described previously[29]. All analyses were performed by investigators blind to experimental conditions. All animal procedures were performed in accordance with institutional guidelines.

**Full Methods** and any associated references are available in the online version of the paper.

**Received 10 November 2011; accepted 11 June 2012.**

**Published online 29 July 2012.**

1. Zhao, C., Deng, W. & Gage, F. H. Mechanisms and functional implications of adult neurogenesis. *Cell* **132,** 645–660 (2008).
2. Kriegstein, A. & Alvarez-Buylla, A. The glial nature of embryonic and adult neural stem cells. *Annu. Rev. Neurosci.* **32,** 149–184 (2009).
3. Ming, G. L. & Song, H. Adult neurogenesis in the mammalian brain: significant answers and significant questions. *Neuron* **70,** 687–702 (2011).
4. Seri, B., Garcia-Verdugo, J. M., McEwen, B. S. & Alvarez-Buylla, A. Astrocytes give rise to new neurons in the adult mammalian hippocampus. *J. Neurosci.* **21,** 7153–7160 (2001).
5. Lagace, D. C. *et al.* Dynamic contribution of nestin-expressing stem cells to adult neurogenesis. *J. Neurosci.* **27,** 12623–12629 (2007).
6. Imayoshi, I. *et al.* Roles of continuous neurogenesis in the structural and functional integrity of the adult forebrain. *Nature Neurosci.* **11,** 1153–1161 (2008).
7. Encinas, J. M. *et al.* Division-coupled astrocytic differentiation and age-related depletion of neural stem cells in the adult hippocampus. *Cell Stem Cell* **8,** 566–579 (2011).
8. Dranovsky, A. *et al.* Experience dictates stem cell fate in the adult hippocampus. *Neuron* **70,** 908–923 (2011).
9. Bonaguidi, M. A. *et al. In vivo* clonal analysis reveals self-renewing and multipotent adult neural stem cell characteristics. *Cell* **145,** 1142–1155 (2011).
10. Encinas, J. M., Vaahtokari, A. & Enikolopov, G. Fluoxetine targets early progenitor cells in the adult brain. *Proc. Natl Acad. Sci. USA* **103,** 8233–8238 (2006).
11. Farrant, M. & Nusser, Z. Variations on an inhibitory theme: phasic and tonic activation of GABA$_A$ receptors. *Nature Rev. Neurosci.* **6,** 215–229 (2005).
12. Bekkers, J. M. & Stevens, C. F. NMDA and non-NMDA receptors are co-localized at individual excitatory synapses in cultured rat hippocampus. *Nature* **341,** 230–233 (1989).
13. Caraiscos, V. B. *et al.* Tonic inhibition in mouse hippocampal CA1 pyramidal neurons is mediated by $\alpha_5$ subunit-containing $\gamma$-aminobutyric acid type A receptors. *Proc. Natl Acad. Sci. USA* **101,** 3662–3667 (2004).
14. Jagasia, R. *et al.* GABA-cAMP response element-binding protein signaling regulates maturation and survival of newly generated neurons in the adult hippocampus. *J. Neurosci.* **29,** 7966–7977 (2009).
15. Freund, T. F. & Buzsaki, G. Interneurons of the hippocampus. *Hippocampus* **6,** 347–470 (1996).
16. Taniguchi, H. *et al.* A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. *Neuron* **71,** 995–1013 (2011).
17. Cardin, J. A. *et al.* Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature* **459,** 663–667 (2009).
18. Ibi, D. *et al.* Social isolation rearing-induced impairment of the hippocampal neurogenesis is associated with deficits in spatial memory and emotion-related behaviors in juvenile mice. *J. Neurochem.* **105,** 921–932 (2008).
19. Andang, M. *et al.* Histone H2AX-dependent GABA$_A$ receptor regulation of stem cell proliferation. *Nature* **451,** 460–464 (2008).
20. Fernando, R. N. *et al.* Cell cycle restriction by histone H2AX limits proliferation of adult neural stem cells. *Proc. Natl Acad. Sci. USA* **108,** 5837–5842 (2011).
21. Lolova, I. & Davidoff, M. Age-related morphological and morphometrical changes in parvalbumin- and calbindin-immunoreactive neurons in the rat hippocampal formation. *Mech. Ageing Dev.* **66,** 195–211 (1992).
22. Satoh, J., Tabira, T., Sano, M., Nakayama, H. & Tateishi, J. Parvalbumin-immunoreactive neurons in the human central nervous system are decreased in Alzheimer's disease. *Acta Neuropathol.* **81,** 388–395 (1991).
23. Masiulis, I., Yun, S. & Eisch, A. J. The interesting interplay between interneurons and adult hippocampal neurogenesis. *Mol. Neurobiol.* **44,** 287–302 (2011).
24. Knable, M. B., Barci, B. M., Webster, M. J., Meador-Woodruff, J. & Torrey, E. F. Molecular abnormalities of the hippocampus in severe psychiatric illness:

postmortem findings from the Stanley Neuropathology Consortium. *Mol. Psychiatry* **9,** 609–620 (2004).

25. Gonzalez-Burgos, G., Fish, K. N. & Lewis, D. A. GABA neuron alterations, cortical circuit dysfunction and cognitive deficits in schizophrenia. *Neural Plast.* **2011,** 723184 (2011).
26. Andre, V., Marescaux, C., Nehlig, A. & Fritschy, J. M. Alterations of hippocampal GABAergic system contribute to development of spontaneous recurrent seizures in the rat lithium-pilocarpine model of temporal lobe epilepsy. *Hippocampus* **11,** 452–468 (2001).
27. Schweizer, C. *et al.* The $\gamma_2$ subunit of $GABA_A$ receptors is required for maintenance of receptors at mature synapses. *Mol. Cell. Neurosci.* **24,** 442–450 (2003).
28. Ge, S. *et al.* GABA regulates synaptic integration of newly generated neurons in the adult brain. *Nature* **439,** 589–593 (2006).
29. Kim, J. Y. *et al.* DISC1 regulates new neuron development in the adult brain via modulation of AKT-mTOR signaling through KIAA1212. *Neuron* **63,** 761–773 (2009).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** We thank L. H. Tsai for initial help in the study; members of the Song and Ming laboratories for discussion; H. Davoudi for help; and Q. Hussaini, Y. Cai and L. Liu for technical support. This work was supported by grants from the National Institutes of Health (NIH) (NS047344) to H.S., the NIH (NS048271, HD069184), the National Alliance for Research on Schizophrenia and Depression and the Adelson Medical Research Foundation to G.L.M., the NIH (MH089111) to B.L., the NIH (AG040209) and New York State Stem Cell Science and the Ellison Medical Foundation to G.E., and by postdoctoral fellowships from the Life Sciences Research Foundation to J.S. and from the Maryland Stem Cell Research Fund to J.S., C.Z. and K.C.

**Author Contributions** J.S. led the project and contributed to all aspects. C.Z., M.A.B., G.J.S., D.H. and K.C. helped with some experiments. Y.G. and S.G. contributed reagents. J.H. provided *SST-Cre* mice. G.E. provided *nestin–GFP* mice. K.D. and K.M. provided initial help on optogenetic tools. B.L. provided $\gamma_2^{f/f}$ mice. J.S., G-l.M. and H.S. designed experiments and wrote the paper.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to H.S. (shongju1@jhmi.edu) or G.M. (gming1@jhmi.edu).

doi:10.1038/nature11303

# Heterodimeric JAK–STAT activation as a mechanism of persistence to JAK2 inhibitor therapy

Priya Koppikar[1]*, Neha Bhagwat[1,2]*, Outi Kilpivaara[1]*, Taghi Manshouri[3], Mazhar Adli[4], Todd Hricik[1], Fan Liu[5], Lindsay M. Saunders[1,2], Ann Mullally[6], Omar Abdel-Wahab[1,7], Laura Leung[1], Abby Weinstein[1], Sachie Marubayashi[1], Aviva Goel[1], Mithat Gönen[8], Zeev Estrov[3], Benjamin L. Ebert[6], Gabriela Chiosis[5], Stephen D. Nimer[5,7], Bradley E. Bernstein[4], Srdan Verstovsek[3] & Ross L. Levine[1,7]

**The identification of somatic activating mutations in *JAK2* (refs 1–4) and in the thrombopoietin receptor gene (*MPL*)[5] in most patients with myeloproliferative neoplasm (MPN) led to the clinical development of JAK2 kinase inhibitors[6,7]. JAK2 inhibitor therapy improves MPN-associated splenomegaly and systemic symptoms but does not significantly decrease or eliminate the MPN clone in most patients with MPN. We therefore sought to characterize mechanisms by which MPN cells persist despite chronic inhibition of JAK2. Here we show that JAK2 inhibitor persistence is associated with reactivation of JAK–STAT signalling and with heterodimerization between activated JAK2 and JAK1 or TYK2, consistent with activation of JAK2 *in trans* by other JAK kinases. Further, this phenomenon is reversible: JAK2 inhibitor withdrawal is associated with resensitization to JAK2 kinase inhibitors and with reversible changes in JAK2 expression. We saw increased JAK2 heterodimerization and sustained JAK2 activation in cell lines, in murine models and in patients treated with JAK2 inhibitors. RNA interference and pharmacological studies show that JAK2-inhibitor-persistent cells remain dependent on JAK2 protein expression. Consequently, therapies that result in JAK2 degradation retain efficacy in persistent cells and may provide additional benefit to patients with JAK2-dependent malignancies treated with JAK2 inhibitors.**

The development of targeted therapies has improved outcomes for patients with kinase-mutant malignancies[8–11]; however, acquired resistance due to mutations in the target kinase[12–14] or in other pathways that render cancer cells insensitive to kinase inhibitor therapy[15–18] remain important clinical concerns. Although JAK inhibitors are now being used to treat patients with MPN, so far JAK inhibitor treatment has not been associated with significant decreases in disease burden in most patients with MPN[6,7]. To understand mechanisms by which MPN cells survive despite chronic JAK kinase inhibition, we performed saturation mutagenesis[19] and next-generation sequencing in cells exposed to two structurally different JAK2 inhibitors, INCB18424 and JAK Inhibitor I. We identified second-site mutations in less than 30–50% of cells exposed to JAK2 inhibitors (Supplementary Table 1). Full-length resequencing of clones proliferating in the presence of INCB18424 or JAK Inhibitor I confirmed the absence of second-site *JAK2* mutations in most surviving clones, and we did not identify second-site *JAK2* mutations in granulocytes from five patients with MPN who had been treated with INCB18424. By contrast, control experiments with mutagenized BCR-Abl cells exposed to imatinib identified more than 20 known, clinically relevant, imatinib resistance alleles[19,20] (data not shown).

These data and clinical experience suggest that the failure of JAK2 inhibitors to decrease disease burden is not due to acquired drug resistance but rather due to persistent growth and signalling in the setting of chronic JAK2 kinase inhibition. We therefore investigated the basis by which JAK2-dependent cells persist despite chronic JAK2 kinase inhibition. We cultured SET-2/UKE-1 (*JAK2V617F*-positive leukaemia) cells and Ba/F3 cells expressing JAK2V617F (EporVF) or MPLW515L (WL) cells with INCB18424 or JAK Inhibitor I for 4–6 weeks. In each case we found that JAK2/MPL-mutant cells could survive and proliferate at inhibitor concentrations sufficient to prevent the growth of parental cells (Fig. 1a, b and Supplementary Figs 1a and 2a). JAK2-inhibitor-persistent ($\text{JAK2}^{\text{Per}}$) cells were resistant to INCB18424-induced apoptosis (Supplementary Fig. 3). *JAK2* resequencing confirmed the absence of second-site mutations in all $\text{JAK2}^{\text{Per}}$ cell lines. $\text{JAK2}^{\text{Per}}$ cells were also insensitive to structurally divergent JAK inhibitors, including TG101348, a JAK2-selective inhibitor in late-stage clinical trials (Fig. 1c and Supplementary Figs 1b, c, 2b and 4). These data indicate that $\text{JAK2}^{\text{Per}}$ cells are

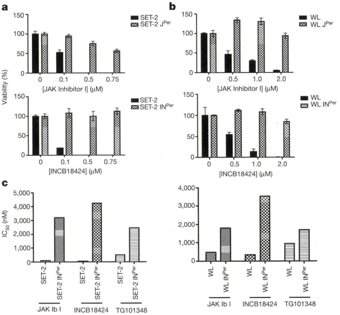


**Figure 1 | Generation of JAK2-inhibitor-persistent cells. a**, **b**, Proliferation of naive and persistent SET-2 (**a**) and WL (**b**) cells with JAK2 inhibitors. Data (means ± s.d.) are from wells plated in triplicate and are representative of three independent experiments. **c**, $\text{IC}_{50}$ values of SET-2 $\text{IN}^{\text{Per}}$ and WL $\text{IN}^{\text{Per}}$ cells exposed to INCB18424, TG101348 and JAK Inhibitor I (JAK Ib I).

---

[1]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [2]Gerstner Sloan- Kettering Graduate School in Biomedical Sciences, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. [3]M. D. Anderson Cancer Center, Houston, Texas 77030, USA. [4]Howard Hughes Medical Institute, Department of Pathology, Massachusetts General Hospital and Broad Institute of Harvard, Massachusetts Institute of Technology, Massachusetts 02114, USA. [5]Molecular Pharmacology and Chemistry, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. [6]Division of Hematology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. [7]Leukemia Service, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. [8]Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA.
*These authors contributed equally to this work.

insensitive to different JAK inhibitors regardless of previous exposure to that inhibitor.

These data are consistent either with the selection of a subpopulation of pre-existing, persistent cells, as previously posited for epidermal growth factor receptor (EGFR) inhibitor-insensitive ‘drug-tolerant persisters’[21], or with the acquisition of persistence by naive, inhibitor-sensitive cells. To distinguish between these possibilities, we derived single-cell clones of inhibitor-naive JAK2/MPL mutant cell lines. Each clonally derived naive cell line was sensitive to JAK inhibitors and retained the capacity to become persistent over time to different JAK inhibitors (Supplementary Fig. 5 and data not shown). These data depict a general capacity for persistence in the absence of clonal selection.

Next, we assessed signalling downstream of JAK2 in JAK2$^{\text{Per}}$ cells. We observed dose-dependent inhibition of downstream signalling in naive cells treated with INCB18424 or JAK Inhibitor I, but not in INCB18424$^{\text{Per}}$ (Fig. 2a and Supplementary Fig. 6a) or JAK Inhibitor I$^{\text{Per}}$ cells (Supplementary Fig. 6b). Similarly, *ex vivo* treatment of granulocytes from patients chronically treated with INCB18424 demonstrated sustained downstream signalling at inhibitor concentrations that inhibited signalling in naive MPN patient samples (Fig. 2b). We then examined whether persistence was associated with constitutive JAK2 activation. We observed persistent phosphorylation of JAK2 in JAK2$^{\text{Per}}$ cells (Supplementary Figs 2c and 6c). Further, gene expression analysis showed that the expression of known JAK–STAT target genes was maintained in JAK$^{\text{Per}}$ cells, whereas these genes were suppressed with acute treatment of inhibitor-naive parental cells (Supplementary Fig. 7).

Given that JAK inhibitors should inhibit JAK2 autophosphorylation, we reasoned that other kinases might associate with and phosphorylate JAK2 in persistent cells. Although EpoR and MPL predominantly signal through JAK2 (ref. 22), previous studies have shown that many cytokine receptors signal through JAK kinase heterodimers[23]. We therefore assessed the activation status of JAK1, JAK3 and TYK2 in naive and persistent SET-2 and WL cells. We observed increased phosphorylation of JAK1 in JAK2$^{\text{Per}}$ cells in comparison with parental cells, whereas TYK2 was constitutively phosphorylated in both parental and JAK2$^{\text{Per}}$ cells (Fig. 2c). Accordingly, immunoprecipitation studies demonstrated that JAK1 and TYK2 associated with phosphoJAK2 in JAK2$^{\text{Per}}$ SET-2, WL (Fig. 2d) and UKE-1 (Supplementary Fig. 2d) cells, but not in the respective parental cells. We saw a similar association between phosphoJAK2 and JAK1 or TYK2 in INCB18424-treated patient samples but not in inhibitor-naive patient samples (Fig. 2e and Supplementary Table 2).

Next, we examined whether the JAK$^{\text{Per}}$ cells were insensitive to JAK inhibitors. *In vitro* kinase assays revealed that the JAK$^{\text{Per}}$ heterodimer complex could phosphorylate myelin basic protein at concentrations of INCB18424 sufficient to inhibit JAK2 kinase activity in naive SET-2 cells (Supplementary Fig. 8). These data suggest that the heterodimer complex in JAK$^{\text{Per}}$ cells retains kinase activity that is relatively insensitive to JAK inhibitors. To determine whether JAK1-mediated phosphorylation of JAK2 was insensitive to INCB18424, we co-expressed a constitutively active mutant form of JAK1 (JAK1V658F)[24] with kinase-dead JAK2 (JAK2K882E) in JAK2-deficient $\gamma$2A cells. We observed persistent JAK2 phosphorylation in JAK1V658F/JAK2K882E $\gamma$2A cells exposed to INCB18424 at concentrations sufficient to inhibit JAK2 autophosphorylation (Supplementary Fig. 9).

We then investigated whether persistence of JAK2 inhibitor was reversible. We removed INCB18424 or JAK Inhibitor I for 2–4 weeks; this led to JAK inhibitor resensitization (Fig. 3a and Supplementary Fig. 10a, b). Resensitized (JAK2$^{\text{Resens}}$) cells were sensitive to all three JAK inhibitors, suggesting that patients with MPN may respond to retreatment or to a different JAK2 inhibitor after a brief withdrawal of treatment. JAK1 or TYK2 association with phosphoJAK2 was lost in JAK2$^{\text{Resens}}$ cells (Fig. 3b and Supplementary Fig. 10c), and activated JAK2 levels were lower in JAK2$^{\text{Resens}}$ cells (Supplementary Fig. 10d).

Previous work attributed persistence in EGFR inhibitor-insensitive ‘drug-tolerant persisters’[21] to the engagement of alternative survival pathways. By contrast, JAK$^{\text{Per}}$ cells were characterized by JAK–STAT pathway reactivation (Fig. 2). We therefore speculated that changes in

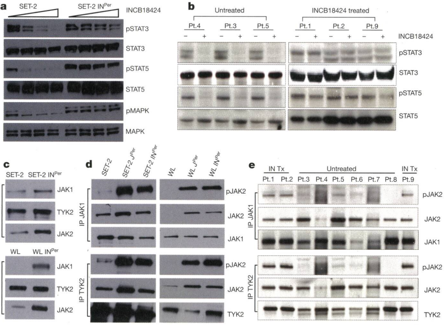


**Figure 2 | Inhibitor-persistent cells and granulocytes from INCB18424-treated patients show continual JAK–STAT signalling and JAK2 activation through transphosphorylation by JAK1 and TYK2. a**, SET-2 and SET-2 IN$^{\text{Per}}$ cells were washed and incubated for 4 h with increasing concentrations of INCB18424 and western blotted. MAPK, mitogen-activated protein kinase. **b**, Granulocytes from naive and INCB18424-treated patients (Pt.) were incubated *ex vivo* for 6 h with dimethylsulphoxide (DMSO) or 150 nM INCB18424 and western blotted. **c**, Increased phosphorylation of JAK1 in persistent cells and constitutive TYK2 phosphorylation in both naive and persistent cells. **d**, Increased association between phosphoJAK2 and both JAK1 and TYK2 in SET-2 JAK$^{\text{Per}}$ cells and increased association between JAK2 and both JAK1 and TYK2 in WL JAK$^{\text{Per}}$ cells. **e**, JAK1/TYK2 association with phosphoJAK2 in granulocytes from three INCB18424-treated (IN Tx) patients, which is not observed in INCB18424-naive MPN samples.

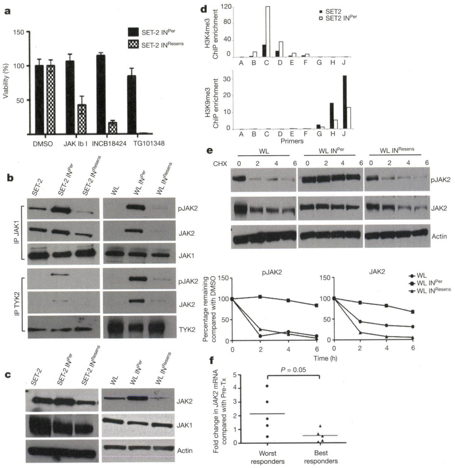


**Figure 3 | JAK2 inhibitor persistence is reversible and JAK2 levels correlate with persistence and resensitization.** **a**, Percentage viability of SET-2 persistent (IN$^{\mathrm{Per}}$) and resensitized (IN$^{\mathrm{Resens}}$) cells at 0.25 μM JAK Inhibitor I, 0.25 μM INCB18424 and 2 μM TG101348. Data (means ± s.d.) are from wells plated in triplicate and are representative of three independent experiments. **b**, Loss of JAK1/TYK2 association with phosphoJAK2 in SET-2 and WL IN$^{\mathrm{Resens}}$ cells. **c**, Reversible changes in JAK2 levels in IN$^{\mathrm{Per}}$ cells compared with naive and IN$^{\mathrm{Resens}}$ SET-2 and WL cells. **d**, ChIP-qPCR of the *JAK2* locus shows increased H3K4me3 and decreased H3K9me3 marks in SET-2 IN$^{\mathrm{Per}}$ cells. **e**, PhosphoJAK2 and total JAK2 levels are degraded on treatment with cycloheximide (CHX; 500 μg ml$^{-1}$ for 2, 4 and 6 h) in naive and resensitized WL cells, but not in IN$^{\mathrm{Per}}$ cells. **f**, Higher *JAK2* levels in INCB18424-treated MPN granulocytes by qRT–PCR compared with those in a small cohort of best responders.

the epigenetic regulation of JAK2 might contribute to JAK inhibitor persistence. *JAK2* messenger RNA (Supplementary Fig. 11) and JAK2 protein (Fig. 3c and Supplementary Figs 2e and 10e) levels were higher in JAK2$^{\mathrm{Per}}$ cells than in parental cells, and were lower in JAK2$^{\mathrm{Resens}}$ cells. Chromatin immunoprecipitation sequencing (ChIP-Seq) analysis of naive *JAK2*-mutant SET-2 cells (M.A., O.A.W., B.E.B. and R.L.L., unpublished observations) revealed that the *JAK2* locus is characterized by trimethylation of histone H3 on Lys 4 (H3K4me3), a modification associated with active promoters, and by H3K9 trimethylation, a mark more typically associated with inactive heterochromatin (Supplementary Fig. 12a and Supplementary Table 3). Analysis of the *JAK2* locus by ChIP coupled to quantitative polymerase chain reaction (ChIP-qPCR) showed a significant increase in H3K4me3 and a decrease in H3K9me3 in JAK2$^{\mathrm{Per}}$ cells in comparison with parental cells (Fig. 3d), which is consistent with a change to a more active chromatin state at the *JAK2* locus. However, global H3K4me3 levels in naive and persistent cells remained unchanged, which is consistent with specific effects on H3K4me3 at the *JAK2* locus in persistent cells (Supplementary Fig. 12b).

Given that JAK2 protein levels, and particularly phosphoJAK2 levels, increased with persistence, we examined whether JAK2 inhibitor persistence was also associated with post-transcriptional stabilization of total and activated JAK2. We have previously shown that JAK2 levels decline rapidly on treatment with cycloheximide in JAK2-mutant cells[25]. We noted a time-dependent decrease in phosphoJAK2 and total JAK2 levels in naive and resensitized WL/SET-2 cells; however, exposure to cycloheximide did not result in a significant decline in JAK2, or more notably in phosphoJAK2, in INCB18424$^{\mathrm{Per}}$ cells (Fig. 3e and Supplementary Fig. 13). These data suggest that chronic treatment with inhibitor results in the stabilization of activated JAK2, which, combined with increased *JAK2* mRNA expression, facilitates the formation of heterodimers.

We then assessed whether this phenomenon was observed *in vivo*. We treated mice engrafted with MPLW515L-mutant murine bone marrow[26] with vehicle or with INCB18424. Treatment with INCB18424 was associated with decreased splenomegaly; however, the proportion of malignant cells was not decreased on treatment with JAK inhibitor, in concordance with our previous results (Supplementary Fig. 14a)[26]. Treatment with INCB18424 was associated with an increase in *JAK2* mRNA and JAK2 protein expression (Supplementary Fig. 14b), similar to that observed in JAK2$^{\mathrm{Per}}$ cells. We also observed an increase in *JAK2* granulocyte mRNA levels in INCB18424-treated patients without clinical or molecular responses, in contrast with patients with clinical or molecular responses to INCB18424 ($P = 0.05$) (Fig. 3f and Supplementary Table 2). Finally, we noted increased JAK2 phosphorylation and increased association between JAK1 and JAK2 in haematopoietic cells from MPLW515L-mutant mice treated with INCB18424. (Supplementary Fig. 14c, d), which is consonant with the expression data.

We examined whether JAK2$^{\mathrm{Per}}$ cells remain JAK2 dependent. JAK2 silencing inhibited proliferation (Fig. 4a), JAK2 activation and

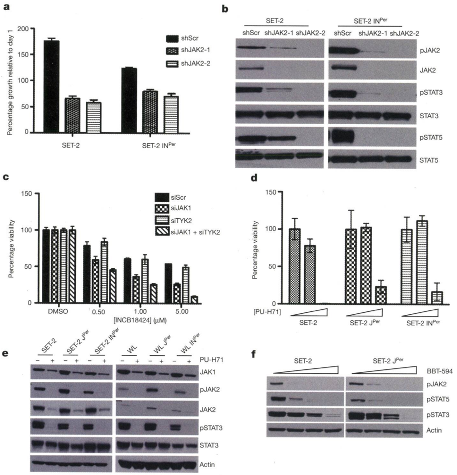


**Figure 4 | Transphosphorylation of JAK2 by JAK1/TYK2 contributes to persistence, and persistent cells can be targeted with type II JAK2 inhibitors or Hsp90 inhibition. a**, SET-2 cells were transfected with non-targeting (shScr) or two JAK2 shRNAs (shJAK2-1 and shJAK2- 2). Viability after 10 days of puromycin selection relative to cell numbers on day 1 is shown. Results are from three biological replicates (means ± s.e.m.). **b**, JAK2 knockdown inhibits signalling in puromycin-selected sensitive and persistent SET-2 cells. **c**, IN$^{\text{Per}}$ SET-2 cells were partly resensitized to INCB18424 after loss of JAK1 or JAK1 + TYK2 using siRNA. Data (means ± s.d.) are from wells plated in triplicate and are representative of three independent experiments. **d**, Naive and persistent SET-2 cells are inhibited by PU-H71. Data (means ± s.d.) are from wells plated in triplicate and are representative of three independent experiments. **e**, PU-H71 degrades JAK2 and inhibits signalling in SET-2 cells. Cells were treated with DMSO or 2 μM PU-H71 (SET-2) and 1 μM PU-H71 (WL) for 16 h. **f**, Treatment with BBT-594 for 4 h inhibits signalling in naive and persistent SET-2 cells.

downstream signalling (Fig. 4b) in naive and JAK2$^{\text{Per}}$ SET-2 cells, which is consistent with a requirement for JAK2 expression in JAK2$^{\text{Per}}$ cells. These data are consistent with previous studies in prolactin receptor cellular systems demonstrating that catalytically inactive JAK2 can serve as a scaffold for transactivation and downstream signalling[27]. However, this had not previously been implicated in JAK-dependent malignancies or in the response to JAK kinase inhibitors. Knockdown of JAK1 and TYK2 increased the sensitivity of SET-2 INCB18424$^{\text{Per}}$ and SET-2 JAK Inhibitor I$^{\text{Per}}$ cells to INCB18424 and JAK Inhibitor I, respectively (Fig. 4c and Supplementary Fig. 15a–c), whereas the parental cells remained unaffected by JAK1 and TYK2 knockdown (Supplementary Fig. 15d). Further, JAK1 and TYK2 knockdown led to decreased downstream signalling and decreased JAK2 phosphorylation in the persistent cells (Supplementary Fig. 15e, f).

We next assessed whether new therapeutic approaches might reverse JAK inhibitor persistence. We previously reported that Hsp90 inhibitors increase JAK2 degradation *in vitro* and *in vivo*[25]. JAK2$^{\text{Per}}$ and parental cells were equally sensitive to Hsp90 inhibition by PU-H71 (Fig. 4d and Supplementary Fig. 16a), and PU-H71 treatment led to JAK2 degradation and inhibited signalling in JAK2$^{\text{Per}}$ cells (Fig. 4e). The currently available type I JAK inhibitors are conformation dependent and can only engage activated JAK2 (ref. 28). We therefore tested the effects of BBT-594, a type II inhibitor that retains the ability to bind inactive JAK2 (ref. 28), in JAK2$^{\text{Per}}$ cells. BBT-594 inhibited the proliferation, JAK activation, and signalling of naive and JAK$^{\text{Per}}$ cells to a similar extent (Fig. 4f and Supplementary Fig. 16b, c).

Taken together, our results suggest that kinase inhibitor persistence can occur through reversible changes in JAK2 expression and transphosphorylation (Supplementary Fig. 17). We show that persistent JAK2 activation in the setting of exposure to JAK inhibitor allows cells to survive without decreasing dependence on JAK2 expression. Consequently, treatments that lead to JAK2 degradation (Hsp90 inhibitors or histone deacetylase inhibitors)[29,30] or that retain the ability to inhibit JAK2 in persistent cells have the potential to improve therapeutic efficacy in patients with MPN.

## METHODS SUMMARY

**Generation of JAK2-inhibitor-persistent cells.** Cells were cultured continuously in increasing concentrations of INCB18424 or JAK Inhibitor I for 4–6 weeks. Cells were considered resistant when the half-maximal inhibitory concentrations (IC$_{50}$ values) of the persistent derivatives were at least double the IC$_{50}$ of parental cells (verified by *in vitro* inhibitor assays). Persistent cells were cultured continuously in the presence of the JAK2 inhibitor. For resensitization experiments, inhibitor was withdrawn from the medium and cells were cultured in the absence of the drug for 2–4 weeks.

**Knockdown of JAK2 and TYK2 in human cell lines.** Short hairpin RNA (shRNA) for JAK2 was purchased from the High Throughput Drug Screening Facility at Memorial Sloan-Kettering Cancer Center, or was a gift from L. Staudt.

shRNA against TYK2 was a gift from T. Look. Whenever required, shRNA oligonucleotides were cloned into pLKO lentiviral systems. Cell lines were transfected with lentivirus, and selected with puromycin. Short interfering RNA (siRNA) targeting either JAK1 or TYK2 was purchased from Invitrogen and used in accordance with the manufacturer's instructions.

**Murine model and analysis of mice.** The MPLW515L murine BMT assay was performed as described previously[5]. Sick mice were randomized to receive INCB18424 twice daily at 60 and 90 mg kg$^{-1}$ or vehicle (0.5% methylcellulose) by oral gavage. Mice were treated for 28 days or until any one of several criteria for killing were met, including moribundity, more than 10% body weight loss, and palpable splenomegaly extending across the midline. Animal care was in strict compliance with Memorial Sloan-Kettering Cancer Center guidelines. Bone marrow and spleen cells were strained and viably frozen in 90% FCS and 10% DMSO.

**Full Methods** and any associated references are available in the online version of the paper.

**Received 11 May 2011; accepted 8 June 2012.**

**Published online 22 July 2012.**

1. James, C. *et al.* A unique clonal *JAK2* mutation leading to constitutive signalling causes polycythaemia vera. *Nature* **434,** 1144–1148 (2005).
2. Kralovics, R. *et al.* A gain-of-function mutation of *JAK2* in myeloproliferative disorders. *N. Engl. J. Med.* **352,** 1779–1790 (2005).
3. Baxter, E. J. *et al.* Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet* **365,** 1054–1061 (2005).
4. Zhao, R. *et al.* Identification of an acquired JAK2 mutation in polycythemia vera. *J. Biol. Chem.* **280,** 22788–22792 (2005).
5. Pikman, Y. *et al. MPLW515L* is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia. *PLoS Med.* **3,** e270 (2006).
6. Verstovsek, S. *et al.* Safety and efficacy of INCB018424, a JAK1 and JAK2 inhibitor, in myelofibrosis. *N. Engl. J. Med.* **363,** 1117–1127 (2010).
7. Pardanani, A. *et al.* Safety and efficacy of TG101348, a selective JAK2 inhibitor, in myelofibrosis. *J. Clin. Oncol.* **29,** 789–796 (2011).
8. Druker, B. J. *et al.* Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N. Engl. J. Med.* **344,** 1031–1037 (2001).
9. Flaherty, K. T. *et al.* Inhibition of mutated, activated BRAF in metastatic melanoma. *N. Engl. J. Med.* **363,** 809–819 (2010).
10. Rosell, R. *et al.* Screening for epidermal growth factor receptor mutations in lung cancer. *N. Engl. J. Med.* **361,** 958–967 (2009).
11. Mok, T. S. *et al.* Gefitinib or carboplatin–paclitaxel in pulmonary adenocarcinoma. *N. Engl. J. Med.* **361,** 947–957 (2009).
12. Kobayashi, S. *et al.* EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **352,** 786–792 (2005).
13. Pao, W. *et al.* Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med.* **2,** e73 (2005).
14. Gorre, M. E. *et al.* Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. *Science* **293,** 876–880 (2001).
15. Engelman, J. A. *et al.* MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science* **316,** 1039–1043 (2007).
16. Pao, W. *et al.* KRAS mutations and primary resistance of lung adenocarcinomas to gefitinib or erlotinib. *PLoS Med.* **2,** e17 (2005).
17. Johannessen, C. M. *et al.* COT drives resistance to RAF inhibition through MAP kinase pathway reactivation. *Nature* **468,** 968–972 (2010).
18. Nazarian, R. *et al.* Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature* **468,** 973–977 (2010).
19. Azam, M., Latek, R. R. & Daley, G. Q. Mechanisms of autoinhibition and STI-571/imatinib resistance revealed by mutagenesis of *BCR-ABL*. *Cell* **112,** 831–843 (2003).
20. Shah, N. P. *et al.* Multiple *BCR-ABL* kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. *Cancer Cell* **2,** 117–125 (2002).
21. Sharma, S. V. *et al.* A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* **141,** 69–80 (2010).
22. Parganas, E. *et al.* Jak2 is essential for signaling through a variety of cytokine receptors. *Cell* **93,** 385–395 (1998).
23. Ihle, J. N. & Gilliland, D. G. Jak2: normal function and role in hematopoietic disorders. *Curr. Opin. Genet. Dev.* **17,** 8–14 (2007).
24. Mullighan, C. G. *et al.* JAK mutations in high-risk childhood acute lymphoblastic leukemia. *Proc. Natl Acad. Sci. USA* **106,** 9414–9418 (2009).
25. Marubayashi, S. *et al.* HSP90 is a therapeutic target in JAK2-dependent myeloproliferative neoplasms in mice and humans. *J. Clin. Invest.* **120,** 3578–3593 (2010).
26. Koppikar, P. *et al.* Efficacy of the JAK2 inhibitor INCB16562 in a murine model of MPLW515L-induced thrombocytosis and myelofibrosis. *Blood* **115,** 2919–2927 (2010).
27. Rider, L., Shatrova, A., Feener, E. P., Webb, L. & Diakonova, M. JAK2 tyrosine kinase phosphorylates PAK1 and regulates PAK1 activity and functions. *J. Biol. Chem.* **282,** 30985–30996 (2007).
28. Andraos, R. *et al.* Modulation of activation-loop phosphorylation by JAK inhibitors is binding mode dependent. *Cancer Discov.* **2,** 512–523 (2012).
29. Wang, Y. *et al.* Cotreatment with panobinostat and JAK2 inhibitor TG101209 attenuates JAK2V617F levels and signaling and exerts synergistic cytotoxic effects against human myeloproliferative neoplastic cells. *Blood* **114,** 5024–5033 (2009).
30. Guerini, V. *et al.* The histone deacetylase inhibitor ITF2357 selectively targets cells bearing mutated JAK2$^{V617F}$. *Leukemia* **22,** 740–747 (2007).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** We thank C. Sawyers, J. Licht, P. Poulikakos and N. Rosen for advice and suggestions; P. Bhatt for his help in the saturation mutagenesis screen; T. Taldone for synthesis of PU-H71; L. Staudt and T. Look for shRNA constructs against JAK2 and TYK2, respectively; and T. Radimerski and P. Manley for providing BBT-594. We are grateful to the Genomics Core Laboratories at Memorial Sloan-Kettering Cancer Center and the Geoffrey Beene Core for their assistance with 454 sequencing. This work was supported in part by National Cancer Institute grant 1R01CA151949-01 to R.L.L., by a grant from the Leukemia and Lymphoma Society to R.L.L. and by a grant from the Myeloproliferative Neoplasms Foundation and the Starr Cancer Consortium to R.L.L., B.E.B. and B.L.E. B.E.B. is a Howard Hughes Medical Institute Early Career Scientist.

**Author Contributions** P.K. and R.L.L. conceived the project. P.K., N.B., O.K. and R.L.L. designed experiments. P.K., N.B., O.K., T.M., M.A., F.L., O.A.W., L.L., A.W., S.M. and A.G. performed experiments. P.K., N.B., T.H., M.G. and M.A. analysed data. L.M.S., A.M., B.L.E. and G.C. provided reagents. Z.E. and S.V. provided patient samples. P.K., N.B. and R.L.L. wrote the paper with input from S.V., Z.E., O.K., B.L.E., B.E.B. and S.D.N.

**Author Information** Microarray data are deposited in the Gene Expression Omnibus under accession number GSE38335. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.L.L. (leviner@mskcc.org).

doi:10.1038/nature11311

# Endogenous antigen tunes the responsiveness of naive B cells but not T cells

Julie Zikherman[1], Ramya Parameswaran[1] & Arthur Weiss[1,2]

**In humans, up to 75% of newly generated B cells and about 30% of mature B cells show some degree of autoreactivity[1]. Yet, how B cells establish and maintain tolerance in the face of autoantigen exposure during and after development is not certain. Studies of model B-cell antigen receptor (BCR) transgenic systems have highlighted the critical role of functional unresponsiveness or 'anergy'[2,3]. Unlike T cells, evidence suggests that receptor editing and anergy, rather than deletion, account for much of B-cell tolerance[4,5]. However, it remains unclear whether the mature diverse B-cell repertoire of mice contains anergic autoreactive B cells, and if so, whether antigen was encountered during or after their development. By taking advantage of a reporter mouse in which BCR signalling rapidly and robustly induces green fluorescent protein expression under the control of the Nur77 regulatory region, antigen-dependent and antigen-independent BCR signalling events *in vivo* during B-cell maturation were visualized. Here we show that B cells encounter antigen during development in the spleen, and that this antigen exposure, in turn, tunes the responsiveness of BCR signalling in B cells at least partly by downmodulating expression of surface IgM but not IgD BCRs, and by modifying basal calcium levels. By contrast, no analogous process occurs in naive mature T cells. Our data demonstrate not only that autoreactive B cells persist in the mature repertoire, but that functional unresponsiveness or anergy exists in the mature B-cell repertoire along a continuum, a fact that has long been suspected, but never yet shown. These results have important implications for understanding how tolerance in T and B cells is differently imposed, and how these processes might go awry in disease.**

A new reporter of antigen receptor signalling was generated recently to examine developmental checkpoints during thymic development[6]. This took advantage of the dynamic expression pattern of the orphan nuclear hormone receptor Nur77 (also known as NR4A1), which is induced rapidly in response to negative selection and T-cell receptor (TCR) stimulation, to develop a green fluorescent protein (GFP) reporter bacterial artificial chromosome (BAC) transgenic line of mice[7]. Interestingly, *Nur77* is also an immediate early gene that is rapidly transcriptionally upregulated in response to BCR signalling[8]. To visualize antigen receptor signalling *in vivo*, we obtained independently generated reporter mice from the Gene Expression Nervous System Atlas (GENSAT) consortium in which enhanced GFP (EGFP) expression is under the control of the Nur77 regulatory region[9]. The founders harboured two distinct insertion sites driving 'high' or 'low' GFP expression. These were independently backcrossed to the C57BL/6 genetic background, yielding $\text{GFP}^{\text{HI}}$ and $\text{GFP}^{\text{LO}}$ lines.

Basal expression of GFP in peripheral $\text{CD4}^{+}$ and $\text{CD8}^{+}$ T cells was higher in both the $\text{GFP}^{\text{HI}}$ and $\text{GFP}^{\text{LO}}$ lines compared to the reporter line described in ref. 6 (Supplementary Fig. 1a). Although basal GFP expression in B cells was substantially higher in the $\text{GFP}^{\text{HI}}$ line relative to the reporter used in ref. 6, the $\text{GFP}^{\text{LO}}$ line failed to express GFP in B cells, suggesting an isolated positional effect. For this reason, all subsequent B-cell studies have focused on the $\text{GFP}^{\text{HI}}$ reporter. After stimulation of thymocytes and peripheral T cells with phorbol myristate acetate (PMA) and/or ionomycin, GFP expression was rapidly induced (Supplementary Fig. 1b; data not shown). *In vitro* stimulation of either the TCR with anti-CD3 or the BCR with anti-IgM also induced GFP expression in a dose-dependent manner (Fig. 1a and Supplementary Fig. 1c; data not shown). $\text{GFP}^{\text{HI}}$ mice were crossed to the IgHEL BCR transgenic line (MD4; in which the immunoglobulin (Ig) receptor specifically recognizes hen egg lysozyme (HEL)) to generate mice with a monoclonal BCR repertoire. The resulting MD4–GFP mice showed dose-dependent GFP induction after treatment with HEL *in vitro* (Fig. 1b and Supplementary Fig. 1d).

To define which antigen-receptor-induced biochemical pathways were required to drive GFP expression, we treated anti-CD3- and anti-IgM-stimulated lymphocytes with a range of small-molecule inhibitors *in vitro*. These experiments showed a nearly complete dependence on Src family kinases in T cells and Syk kinase in B cells

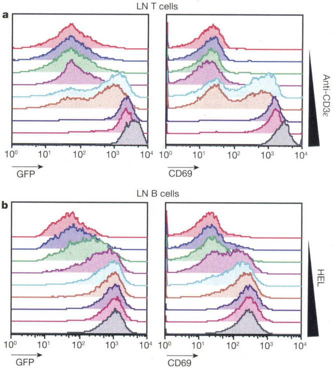


**Figure 1 | The Nur77–GFP BAC transgenic reporter is responsive to antigen-receptor signalling *in vitro*.** **a**, Histograms represent GFP and CD69 expression of $\text{GFP}^{\text{LO}}$ transgenic lymph node (LN) T cells treated with varying doses of plate-bound anti-CD3ε for 16 h (0.00625–6.4 μg ml$^{-1}$ in a fourfold dilution series). **b**, Histograms represent GFP and CD69 expression in IgHEL $\text{GFP}^{\text{HI}}$ transgenic lymph node B cells treated with varying doses of HEL for 16 h (0.125–16 ng ml$^{-1}$ in a twofold dilution series). Data are representative of at least three independent experiments.

---

[1]Division of Rheumatology, Department of Medicine, Rosalind Russell Medical Research Center for Arthritis, University of California, San Francisco, California 94143, USA. [2]Howard Hughes Medical Institute, University of California, San Francisco, California 94143, USA.

(Supplementary Fig. 1e, f). In B cells, GFP expression was partially dependent on the protein kinase C (PKC), calcineurin, mitogen-activated protein kinase (MAPK) and phosphatidylinositol-3-OH kinase (PI(3)K) pathways, whereas in T cells, GFP expression most clearly required PKC (Supplementary Fig. 1e, f).

To define whether signals other than antigen-receptor ligation were sufficient to drive GFP expression in B cells, we treated GFP$^{HI}$ B cells *in vitro* with various stimuli. Toll-like receptor (TLR)-4 and TLR9 ligands, along with anti-CD40, could drive GFP expression in B cells, but this effect was considerably less robust than anti-IgM stimulation (Supplementary Fig. 1g). Notably, B-cell activating factor (BAFF) treatment with doses as high as 200 ng ml$^{-1}$, sufficient to induce prolonged B-cell survival *in vitro*, failed to induce GFP-reporter expression in B cells (Supplementary Fig. 1g).

The reporter responded to TCR-dependent signalling *in vivo*, as shown by GFP expression at TCR-dependent checkpoints during thymic development. Signalling through the pre-TCR, comprised of a recombined TCR-β chain and the invariant pre-TCR-α chain, drives developing thymocytes to transit the β-selection checkpoint. We observed abrupt upregulation of GFP expression at the 'double-negative' DN3b stage of development, precisely at the β-selection checkpoint transition (Supplementary Fig. 2a).

After successful transit through the β-selection checkpoint, double-negative thymocytes upregulate the CD4 and CD8 coreceptors, and recombine the TCR-α chain to express a mature αβTCR. These cells then undergo TCR-dependent positive or negative selection. We observed marked GFP upregulation in post-selection CD69$^{HI}$ TCR-β$^{HI}$ 'double-positive' thymocytes (Supplementary Fig. 2b), as found in ref. 6.

It has been speculated that, at the border of positive and negative selection, SP4$^{+}$ thymocytes can be rescued from death by adopting the regulatory T-cell (T$_{reg}$) fate. Indeed, CD25$^{+}$ SP4$^{+}$ thymocytes expressed much higher GFP levels than conventional SP4$^{+}$ thymocytes, indicating that strong TCR signalling favours the T$_{reg}$ fate, in agreement with the results from ref. 6 (Supplementary Fig. 2c).

We reported that titration of CD45 expression in an allelic series of mice regulates TCR signalling during thymic development[10]. We crossed the GFP$^{HI}$ reporter onto a genetic background harbouring two copies of the Lightning (L) *CD45* (also known as *Ptprc*) allele, in which a point mutation in the extracellular domain leads to reduced surface expression of CD45 (15% of expression levels in wild-type mice)[10]. Both the fraction of high-GFP-expressing cells and the average GFP content of post-selection double-positive thymocytes was markedly reduced in so-called L/L GFP mice (Supplementary Fig. 2d). This result indicates that the GFP reporter is indeed sensitive to genetic titration of TCR signal strength.

To identify analogous BCR-dependent signalling checkpoints during B-cell development, we assessed successive stages of bone marrow B-cell development in GFP$^{HI}$ reporter mice[11] (Fig. 2a and Supplementary Fig. 3a, b). We observed virtually no GFP expression except in the mature B cells that recirculate to the bone marrow (Hardy Fraction F; IgM$^{LO}$IgD$^{HI}$), indicating that GFP upregulation occurs sometime after the early bone marrow stages of development, despite evidence of the contribution of antigen encounter to deletion and receptor editing in the bone marrow[12].

Splenic B-cell development, which follows maturation in the bone marrow, is subdivided into successive transitional stages[13–15]. We observed a bimodal distribution of GFP expression among splenic B cells and found that early transitional B cells (T1) are largely GFP negative, but that later transitional stages (T2 and T3) contained a large proportion of GFP-positive B cells (Fig. 2b, c). Mature follicular B cells were mostly GFP positive and showed a broad distribution of GFP expression (Fig. 2c). Notably, a similar pattern of GFP expression, albeit at much lower levels, was evident in an independently generated GFP reporter[6] (Supplementary Fig. 3c). GFP expression across these splenic developmental stages inversely correlated with surface IgM expression (Fig. 2d). Transitional B-cell stages have previously been subdivided into T2 and T3 stages on the basis of surface IgM

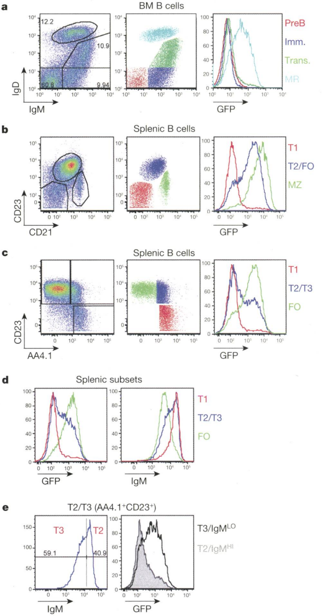


**Figure 2 | Expression of the Nur77–GFP BAC transgenic reporter is upregulated at specific checkpoints during B-cell development. a**, Left, plot of GFP$^{HI}$ transgenic bone marrow (BM) CD19$^{+}$ B cells stained for IgM and IgD to identify pre-B, immature (imm.), transitional (trans.) and mature recirculating (MR) subsets (counter-clockwise from bottom left corner). Middle, bone marrow subsets are colour coded. Right, overlaid histograms representing GFP expression in these subsets. **b**, Left, plot of GFP$^{HI}$ transgenic splenic CD19$^{+}$ B cells stained for CD23 and CD21 expression to identify T1 (CD23$^{-}$ CD21$^{-}$), T2/Follicular (FO; CD23$^{+}$ CD21$^{+}$) and marginal zone (MZ; CD21$^{HI}$) subsets. Middle, splenic B-cell subsets are colour coded. Right, overlaid histograms represent GFP expression in these subsets. **c**, Left, plot of GFP$^{HI}$ transgenic splenic CD19$^{+}$ B cells, excluding marginal zone compartment, stained for CD23 and AA4.1 expression to identify T1 (AA4.1$^{+}$ CD23$^{-}$), T2/3 (AA4.1$^{+}$ CD23$^{+}$) and follicular (AA4.1$^{-}$ CD23$^{+}$) subsets. Middle, splenic B-cell subsets are colour coded. Right, overlaid histograms represent GFP expression in these subsets. **d**, Overlaid histograms represent GFP (left) and IgM (right) expression in T1, T2/3 and follicular subsets as identified in **c**. **e**, Left, T2/3 (AA4.1$^{+}$ CD23$^{+}$) B-cell subset subdivided by IgM expression into T2 (IgM$^{HI}$) and T3 (IgM$^{LO}$) stages. Right, overlaid histograms represent GFP expression T2 and T3 subsets. All data are representative of at least three independent experiments.

downregulation[15] (Fig. 2e). We observed that GFP upregulation seems to occur at precisely this transition between the T2 and T3 stages (Fig. 2e and Supplementary Fig. 3d, e).

Interestingly, *in vitro* BCR stimulation of bone marrow and splenic B-cell subsets resulted in GFP upregulation to differing extents. Minimal GFP upregulation was seen in bone marrow immature and transitional stages, but robust upregulation was evident in splenic T1, T2 and follicular B cells (Supplementary Fig. 4). This indicates that splenic, but not bone marrow, subsets have the capacity to upregulate GFP.

To determine whether the amount of GFP expression in unstimulated B cells reflected BCR signal strength/antigen exposure, we took advantage of our previously characterized allelic series of CD45-expressing mice[10,16]. In these animals, CD45 expression is genetically varied across a broad range and correlates with BCR signal strength[16]. L/L mice with reduced surface expression of CD45 show impaired BCR signal transduction. So-called H/− mice express a normally splicing *CD45* transgene superimposed on endogenous wild-type *CD45* to produce an animal with supraphysiologic CD45 expression. B cells from these mice show enhanced BCR signal strength. After crossing the Nur77–GFP reporter mouse to the CD45 allelic series, we noted that GFP expression at the T1 stage was unaffected, whereas increasing CD45 expression resulted in a higher proportion of GFP-positive B cells at the T2 stage (Fig. 3a and Supplementary Fig. 5a). Notably, the distribution of GFP expression in this compartment remained bimodal, further supporting the notion that a discrete signalling event occurs at this stage, the threshold of which is regulated by CD45 and BCR signal strength. GFP expression in follicular mature B cells was markedly reduced in L/L mice, consistent with a reduction in BCR signal strength, but was minimally altered in H/− mice with higher CD45 expression (Fig. 3a and Supplementary Fig. 5b). However, modulation of GFP expression by CD45 was much more apparent in the marginal zone compartment, suggesting an exquisite sensitivity to BCR signal strength (Fig. 3a and Supplementary Fig. 5b).

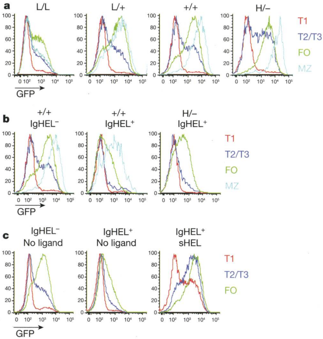


**Figure 3 | Expression of the Nur77–GFP BAC transgenic reporter is sensitive to genetic modulation of BCR signal strength and to antigen.** **a**, CD45 allelic series (low to high CD45 expression: L/L, L/+, +/+ and H/−) GFP$^{HI}$ transgenic splenic B cells were stained to identify B-cell subsets as in Fig. 2b, c. Overlaid histograms represent GFP expression in splenic subsets as gated in Supplementary Fig. 5a. **b**, CD45$^{+/+}$ GFP$^{HI}$ transgenic and H/− GFP$^{HI}$ transgenic splenic B cells with an unrestricted (no IgHEL transgene; IgHEL$^-$) or restricted (IgHEL$^+$) repertoire in the absence of sHEL antigen were analysed as in **a**. Overlaid histograms represent GFP expression in splenic subsets as gated in Supplementary Fig. 5c. **c**, CD45$^{+/+}$ GFP$^{HI}$ transgenic splenic B cells with an unrestricted or restricted repertoire in the presence or absence of sHEL antigen were analysed as in **a**. Overlaid histograms represent GFP expression in splenic subsets as gated in Supplementary Fig. 5d. All animals in these experiments were generated through genetic crosses. All data are representative of at least five independent experiments.

As Nur77–GFP expression is regulated by modulation of BCR signal strength (Fig. 3a and Supplementary Fig. 5b), we proposed that endogenous antigen exposure might drive BCR signalling during maturation of wild-type B cells with a diverse repertoire. To explore this possibility, we took advantage of the IgHEL/soluble (s)HEL double-transgenic system (MD4/ML5), in which MD4 mice with a monoclonal IgHEL BCR can be studied in the presence or absence of sHEL[2]. In the Nur77–GFP reporter mice with the IgHEL BCR transgene-restricted repertoire in the absence of antigen, we observed a marked reduction in GFP in splenic B cells (Fig. 3b and Supplementary Fig. 5b, c). Notably, the bimodal distribution of GFP expression observed in the context of a wild-type repertoire was lost in these mice. Further increasing CD45 expression in the context of such a restricted repertoire to increase tonic BCR signalling resulted in increasing GFP expression, but again only in a unimodal rather than a bimodal distribution (Fig. 3b and Supplementary Fig. 5c). Finally, the introduction of sHEL ligand by crossing ML5 (sHEL trangenic) mice to IgHEL transgenic reporter mice resulted in increased GFP expression as expected, and remarkably reconstituted bimodal GFP expression in the transitional splenic stages of development (Fig. 3c and Supplementary Figs 5d and 6). These data indicate that normal B-cell development is characterized by a wide range of antigen experience, and that Nur77-driven GFP distribution in follicular mature B cells serves as a marker of such exposure.

To determine whether antigen recognition during splenic B-cell development had functional effects on signalling, we selectively gated on the extremes of GFP expression. We observed that high-GFP-expressing B cells had dampened 40s ribosomal protein S6 (RPS6) phosphorylation (a PI(3)K-dependent event) and calcium entry relative to low-GFP-expressing B cells in response to IgM ligation (Supplementary Figs 4a and 7a). Moreover, we observed that basal calcium levels were elevated in high-GFP-expressing B cells, reminiscent of anergic B cells identified in various model BCR transgenic systems[17,18]. Dampened inducible signalling and increased basal calcium were not isolated properties of very-high-GFP-expressing B cells, but rather seemed to represent continuous functional properties across the entire spectrum of GFP expression of mature follicular B cells (Fig. 4a). Furthermore, restricting the BCR repertoire in the absence of ligand ablated differences in functional responsiveness, but not in basal calcium (Supplementary Fig. 7b). Notably, neither inducible calcium responses nor basal calcium levels correlated with GFP expression in naive CD25$^+$ CD4$^+$ T cells, indicating that only in B cells does antigen exposure tune functional responsiveness (Supplementary Fig. 8a).

Mature B cells express two isotypes of the BCR, IgM and IgD. We wanted to determine whether the functional responsiveness in GFP B cells was modulated in response to stimulation through the IgD BCR in the same manner as it is to the IgM BCR. We found that this was not the case (Fig. 4b, c); responsiveness to IgM BCR stimulation was markedly blunted in cells with high GFP expression, whereas IgD responsiveness remained intact. Stimulation with anti-κ antibodies to ligate both surface IgM and IgD resembled isolated IgD stimulation (Supplementary Fig. 8b). By simultaneously staining for surface IgM expression with a nonstimulatory monovalent Fab fragment and assessing calcium responses in GFP B cells, we show that differences in surface IgM expression largely accounted for the functional differences at different levels of GFP expression (Supplementary Fig. 8c). However, basal calcium differences were independent of surface IgM expression (Supplementary Fig. 8d).

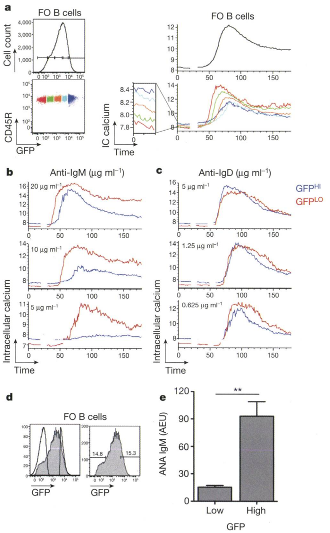


**Figure 4 | GFP expression predicts functional responsiveness and autoreactivity of B cells.** **a**, Left, GFP$^{HI}$ transgenic follicular (CD23$^+$ AA4.1$^-$) splenic B cells were subdivided into colour-coded bins on the basis of GFP expression. Right, GFP$^{HI}$ transgenic splenic B cells were loaded with Indo-1 dye and stimulated with 10 μg ml$^{-1}$ anti-IgM. Ratiometric assessment of intracellular calcium was carried out by flow cytometry. Upper right panel represents calcium entry in total follicular splenic B cells. Lower right panel represents basal and inducible intracellular calcium in GFP-specific bins. IC, intracellular. **b**, **c**, Intracellular calcium entry was assessed in GFP$^{HI}$ transgenic follicular splenic B cells after anti-IgM (**b**) or anti-IgD (**c**) stimulation at varying doses. High and low GFP-expressing gates are overlaid. Data in **a**–**c** are representative of at least three independent experiments. **d**, Overlaid histograms represent pre- and post-sort follicular mature CD23$^+$ AA4.1$^-$ B cells as gated in Supplementary Fig. 9. The 15% lowest and highest GFP fractions from the follicular mature B-cell compartment were selected for sorting. **e**, High-expressing-GFP and low-expressing-GFP B cells sorted as described in **c** and Supplementary Fig. 9a were stimulated *in vitro* with LPS for 4 days. Supernatants were subjected to ANA IgM enzyme-linked immunosorbent assays (ELISA) and total IgM ELISA. The graph represents quantification of ANA IgM normalized to total IgM from four independent sorting experiments. Data are ± s.e.m. AEU, arbitrary ELISA units. Significance was assessed by unpaired *t*-test. $^{**}P < 0.005$.

Additional characteristics of monoclonal BCR transgenic models of anergic B cells include a failure to upregulate activation markers in response to various stimuli[3]. We stimulated sorted high- and low-GFP-expressing B cells and observed that activation marker upregulation in response to IgM stimulation is impaired in high-GFP-expressing B cells (Supplementary Figs 9, 10). Importantly, responses to lipopolysaccharide (LPS) and CD40 were unaffected, as was *in vitro* survival in the presence or absence of BAFF (data not shown).

Finally, to determine directly whether the BCR repertoire of mature B cells with high GFP expression and impaired functional responses was indeed autoreactive, sorted high- and low-GFP-expressing B cells were polyclonally stimulated *in vitro* with LPS, and secreted antibody was assessed for anti-nuclear antibody (ANA) reactivity. (Fig. 4d, e and Supplementary Fig. 9). Notably, neither cell proliferation nor antibody secretion following LPS stimulation were impaired in high-GFP B cells (data not shown). We found a significant increase in ANA reactivity, suggesting auto- or polyreactivity in the repertoire of such naturally occurring anergic B cells (Fig. 4d, e).

The human B-cell repertoire is characterized by a high prevalence of polyreactive and autoreactive BCRs[1,19]. Anergy or functional unresponsiveness may serve to keep such autoreactive clones in check[3]. Array data have shown that wild-type B cells have an intermediate phenotype between antigen-naive and anergic B cells, suggesting the possible presence of anergic B cells in the wild-type mature repertoire[20,21]. It has recently been argued that the so-called T3 splenic subset may in fact represent sequestered anergic B cells rather than an intermediate developmental stage[22,23]. However, the prevalence of anergy in the normal mature B-cell repertoire has not been clear[21]. We show that there is a continuum of anergy or unresponsiveness to anti-IgM stimulation in the mature B-cell compartment, and that this responsiveness is, in turn, tuned by developmental antigen recognition.

It has long been observed that marked IgM downregulation is seen in BCR transgenic systems in the presence of either antigen or enhanced BCR signal strength[2,16,19,24–26]. IgD, by contrast, remains relatively unmodulated in these systems. Here, we show that, in the wild-type B-cell repertoire, IgM downregulation correlates with the extent of antigen recognition during development and accounts for dampened B-cell responses to anti-IgM stimulation, whereas IgD expression and responses are intact. We suggest that this constitutes a general mechanism to modulate BCR signalling in autoreactive B cells, but permits them to persist as a pool of extended antibody specificity for purposes of protective immunity. Indeed, we demonstrate an increased proportion of ANA-reactive BCR specificities in high-GFP-expressing B cells, suggesting that these cells are auto- or polyreactive. It is tempting to speculate that this large reservoir of dormant autoreactive B cells in the mature BCR repertoire may serve as the source of pathogenic autoantibodies that characterize rheumatic diseases such as systemic lupus erythematosus.

## METHODS SUMMARY

The following mouse strains have been previously described: the CD45 allelic series including Lightning (L/L), H/− (HE) mice[10,16,27], IgHEL (MD4) and sHEL (ML5) mice[2]. Nur77–EGFP BAC transgenic mice were obtained from the GENSAT consortium[9]. Nur77–GFP reporter mice described in ref. 6 were supplied by the Hogquist laboratory. All strains were backcrossed to the C57BL/6 genetic background at least six generations and were maintained in the University of California, San Francisco animal facility in accordance with institutional regulations. *In vitro* lymphocyte-stimulation assays were performed as previously described on plates containing either soluble anti-IgM Fab′2, precoated with anti-CD3ε, and/or containing various stimuli and inhibitors[28]. Calcium assays were performed as previously described[28], except that Indo-1 dye (Invitrogen) was used to load cells, and an ultraviolet laser on the BD Fortessa was used for detection. Intracellular phospho-S6 staining and stimulation was performed as previously described[16]. Sorting of GFP-high and -low-expressing B cells using a MoFlo cell sorter was performed as follows: splenic and lymph node cells were pooled and stained to identify DAPI (4′,6-diamidino-2-phenylindole)–CD23$^+$ AA4.1 mature B cells. The highest and lowest 15% of GFP-expressing B cells were retrieved and were incubated with varying stimuli. Sorted cells were plated at a concentration of $1.5 \times 10^6$ cells per ml in complete DMEM media and were stimulated with anti-IgM Fab′2 at varying doses for 16 h to assess activation marker upregulation. Alternatively, sorted cells were incubated with 10 μg ml$^{-1}$ LPS at a concentration of $6 \times 10^6$ cells per ml in complete DMEM to drive polyclonal antibody secretion. Supernatants were then collected and subjected to enzyme-linked immunosorbent assay (ELISA). The assay to detect total IgM was performed as previously described[29]. The ANA ELISA kit obtained from Inova Inc. was used as per manufacturer's instructions.

**Full Methods** and any associated references are available in the online version of the paper.

**Received 9 March 2012; accepted 11 June 2012.**

**Published online 19 August 2012.**

1. Wardemann, H. *et al.* Predominant autoantibody production by early human B cell precursors. *Science* **301,** 1374–1377 (2003).
2. Goodnow, C. C. *et al.* Altered immunoglobulin expression and functional silencing of self-reactive B lymphocytes in transgenic mice. *Nature* **334,** 676–682 (1988).
3. Cambier, J. C., Gauld, S. B., Merrell, K. T. & Vilen, B. J. B-cell anergy: from transgenic models to naturally occurring anergic B cells? *Nature Rev. Immunol.* **7,** 633–643 (2007).
4. Lang, J. *et al.* Enforced Bcl-2 expression inhibits antigen-mediated clonal elimination of peripheral B cells in an antigen dose-dependent manner and promotes receptor editing in autoreactive, immature B cells. *J. Exp. Med.* **186,** 1513–1522 (1997).
5. Halverson, R., Torres, R. & Pelanda, R. Receptor editing is the main mechanism of B cell tolerance toward membrane antigens. *Nature Immunol.* **5,** 645–650 (2004).
6. Moran, A. E. *et al.* T cell receptor signal strength in Treg and iNKT cell development demonstrated by a novel fluorescent reporter mouse. *J. Exp. Med.* **208,** 1279–1289 (2011).
7. Winoto, A. & Littman, D. R. Nuclear hormone receptors in T lymphocytes. *Cell* **109** (Suppl. 1), S57–S66 (2002).
8. Mittelstadt, P. R. & DeFranco, A. L. Induction of early response genes by cross-linking membrane Ig on B lymphocytes. *J. Immunol.* **150,** 4822–4832 (1993).
9. The Gene Expression Nervous System Atlas (GENSAT) Project. NINDS Contract # N01NS02331 to The Rockefeller University http://www.gensat.org/index.html (New York, USA).
10. Zikherman, J. *et al.* CD45–Csk phosphatase–kinase titration uncouples basal and inducible T cell receptor signaling during thymic development. *Immunity* **32,** 342–354 (2010).
11. Hardy, R. R., Carmack, C. E., Shinton, S. A., Kemp, J. D. & Hayakawa, K. Resolution and characterization of pro-B and pre-pro-B cell stages in normal mouse bone marrow. *J. Exp. Med.* **173,** 1213–1225 (1991).
12. Goodnow, C. C., Sprent, J., Fazekas de St Groth, B. & Vinuesa, C. G. Cellular and genetic mechanisms of self tolerance and autoimmunity. *Nature* **435,** 590–597 (2005).
13. Loder, F. *et al.* B cell development in the spleen takes place in discrete steps and is determined by the quality of B cell receptor-derived signals. *J. Exp. Med.* **190,** 75–90 (1999).
14. Chung, J. B., Silverman, M. & Monroe, J. G. Transitional B cells: step by step towards immune competence. *Trends Immunol.* **24,** 342–349 (2003).
15. Allman, D. *et al.* Resolution of three nonproliferative immature splenic B cell subsets reveals multiple selection points during peripheral B cell maturation. *J. Immunol.* **167,** 6834–6840 (2001).
16. Zikherman, J., Doan, K., Parameswaran, R., Raschke, W. & Weiss, A. Quantitative differences in CD45 expression unmask functions for CD45 in B-cell development, tolerance, and survival. *Proc. Natl Acad. Sci. USA* **109,** E3–E12 (2012).
17. Cooke, M. *et al.* Immunoglobulin signal transduction guides the specificity of B cell–T cell interactions and is blocked in tolerant self-reactive B cells. *J. Exp. Med.* **179,** 425–438 (1994).
18. Yarkoni, Y., Getahun, A. & Cambier, J. C. Molecular underpinning of B-cell anergy. *Immunol. Rev.* **237,** 249–263 (2010).
19. Duty, J. A. *et al.* Functional anergy in a subpopulation of naive B cells from healthy humans that express autoreactive immunoglobulin receptors. *J. Exp. Med.* **206,** 139–151 (2009).
20. Glynne, R. *et al.* How self-tolerance and the immunosuppressive drug FK506 prevent B-cell mitogenesis. *Nature* **403,** 672–676 (2000).
21. Glynne, R., Ghandour, G., Rayner, J., Mack, D. H. & Goodnow, C. C. B-lymphocyte quiescence, tolerance and activation as viewed by global gene expression profiling on microarrays. *Immunol. Rev.* **176,** 216–246 (2000).
22. Merrell, K. *et al.* Identification of anergic B cells within a wild-type repertoire. *Immunity* **25,** 953–962 (2006).
23. Teague, B. N. *et al.* Cutting edge: transitional T3 B cells do not give rise to mature B cells, have undergone selection, and are reduced in murine lupus. *J. Immunol.* **178,** 7511–7515 (2007).
24. Cornall, R. J. *et al.* Polygenic autoimmune traits: Lyn, CD22, and SHP-1 are limiting elements of a biochemical pathway regulating BCR signaling and selection. *Immunity* **8,** 497–508 (1998).
25. Benschop, R. J. *et al.* Activation and anergy in bone marrow B cells of a novel immunoglobulin transgenic mouse that is both hapten specific and autoreactive. *Immunity* **14,** 33–43 (2001).
26. Fields, M. L. & Erikson, J. The regulation of lupus-associated autoantibodies: immunoglobulin transgenic models. *Curr. Opin. Immunol.* **15,** 709–717 (2003).
27. Virts, E. L., Diago, O. & Raschke, W. C. A CD45 minigene restores regulated isoform expression and immune function in CD45-deficient mice: therapeutic implications for human CD45-null severe combined immunodeficiency. *Blood* **101,** 849–855 (2003).
28. Zikherman, J. *et al.* *PTPN22* deficiency cooperates with the CD45 E613R allele to break tolerance on a non-autoimmune background. *J. Immunol.* **182,** 4093–4106 (2009).
29. Hermiston, M. L., Tan, A. L., Gupta, V. A., Majeti, R. & Weiss, A. The juxtamembrane wedge negatively regulates CD45 function in B cells. *Immunity* **23,** 635–647 (2005).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** We thank A. Roque for assisting with animal husbandry and Z. Wang and J. Paw for help with cell sorting. This work was supported by the Rosalind Russell Medical Research Foundation Bechtel Award (J.Z.), an American College of Rheumatology REF Rheumatology Investigator Award (J.Z.), an Arthritis National Research Foundation grant (J.Z.) and National Institutes of Health Grant K08 AR059723 (J.Z.), as well as the Howard Hughes Medical Institute (A.W.).

**Author Contributions** J.Z. and A.W. designed the research, J.Z. and R.P. performed the research, J.Z. and R.P. analysed the data and J.Z. and A.W. wrote the manuscript.

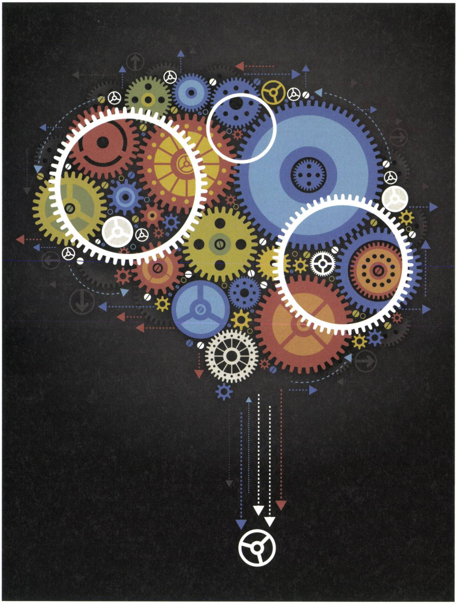
**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.W. (aweiss@medicine.ucsf.edu).

# CAREERS

**EUROPE** Rise in business research spending could prompt recruitment **p.167**

**ENTREPRENEURSHIP** Trusted mentors are best at teaching business skills **p.167**

**NATUREJOBS** For the latest career listings and advice **www.naturejobs.com**



VLADGRIN/SHUTTERSTOCK

**COLLABORATIONS**

# A single cog in a complex machine

*If they remain vigilant, early-career researchers can reap benefits from taking part in big international projects.*

**BY SARAH KELLOGG**

Catherine Luria has little doubt about the benefits of participating in a big, international collaboration. Luria, a marine microbiologist beginning her third year of graduate study at Brown University in Providence, Rhode Island, is examining how changes in sea-ice coverage and blooms of phytoplankton affect bacterial diversity from season to season. She has literally gone to the ends of the Earth to join a collaboration: the Palmer Antarctica Long Term Ecological Research (LTER) project on the western coast of the Antarctic Peninsula.

Luria will return to Antarctica this month, and several more times over the next two years, taking a week to travel there to spend two months with about 25 researchers and another dozen support staff involved in LTER. While there, she will characterize the water column, collect water samples and measure bacterial and phytoplankton abundance and bacterial production in the lab. She will examine microbial growth rates, physiology and community composition under different conditions.

"It's a huge networking opportunity at this stage in my career," says Luria. Thanks to the collaboration, she will be able to work with many more measurements than she would have on her own. "What has proved to be especially helpful is having access to data," she says. "Suddenly I'm able to dip into this pool of high-quality, curated data going back a decade or more. I have the ability to get more meaningful results. It's not data from a snapshot of when your grant just happened to be funded."

High-profile international research projects can bring together hundreds, if not thousands, of scientists. Joining one is no guarantee of professional success for an early-career researcher, but it does provide an exceptional environment for learning, and access to crucial data and networking opportunities that can advance personal research and open professional doors.

Team science practised on a huge scale not only yields ground-breaking results, but can also establish and fortify careers, as researchers have found in ventures such as the Human Genome Project; the ATLAS particle-physics experiment at the Large Hadron Collider at CERN, Europe's particle-physics laboratory ▶

ENCODE
Encyclopedia of DNA Elements
nature.com/encode

near Geneva in Switzerland; or the Encyclopedia of DNA Elements (ENCODE), a project to define the functional elements of the human genome (see page 45).

"Research careers are built inside big collaborations, and a measure of the success of university-based research groups is the number and quality of prominent positions within collaborations that are held by the group's members," says Ricardo Gonçalo, a particle physicist at Royal Holloway, University of London, who has worked on ATLAS.

Positions in big consortia are highly sought after, but drawbacks to participating include limited access to principal investigators; constant jockeying for recognition; the pressure to subjugate personal research to elevate project research; the risk of getting lost in long lists of authors on publications; and the difficulty of distinguishing individual work from group work. Having so many people in a project "implies a lot of politics, different ways of behaviour that affect our interaction, many rules", says Patricia Conde Muíño, a physicist at the Laboratory of Instrumentation and Experimental Particle Physics in Lisbon, who worked on HERA-B, an experiment at the DESY particle accelerator in Hamburg, Germany, that included 32 institutes and 250 collaborators from 13 countries. "One thing that sometimes is complicated is the internal competition. This is stronger in the physics groups, where there are literally hundreds of people trying to do the same thing as you," she adds.

> ***"These are incredibly exciting and important projects, and they're seen as the future of science by some."***

**VALUE ADDED**

Veterans of consortia say that it is crucial for young scientists to consult experienced investigators when considering whether to join a project. They should weigh their research objectives and career goals, and assess how their strengths and weaknesses might be elevated or strained on a high-profile project. Although it is impossible to know how individual graduate students or postdocs will fare in such intense environments, it is important for them to go into projects with their eyes open to potential challenges. Those who don't proactively seek to develop their skills and network with established researchers may end up being little more than Anonymous Author Number 16 on a 40-author publication.

The search for a high-profile collaboration begins most effectively with a review of personal career goals and how best to achieve them. Large consortia often represent just one step on a long career path. Young scientists can use self-assessment tools and resources to look at their core competencies and to evaluate long-term goals to see how a large project could match their aims.

Armed with this knowledge, postdocs should talk to trusted faculty members or mentors, and seek out scientists from the collaboration who are speaking or presenting posters at conferences. These people can alert the young scientist to research opportunities and provide key contacts to enable them to visit labs and meet principal investigators. The aim is to find the project that best fits the young researcher's professional interests and personal circumstances, and networking is the most efficient way to do that (see 'Look before you leap').

Joining a high-profile collaboration opens the door to research and colleagues that may previously have been out of reach, while also providing the rare opportunity to explore cutting-edge research in a competitive and well-funded environment. The intimate collaborations of smaller-group research are lost, but access to international experts gives young researchers great opportunities at this crucial time in their careers.

High-profile projects also give researchers a chance to learn new methods and processes from international colleagues who bring very different approaches to the scientific enterprise. "I think this brings enrichment, and hopefully you're able to pick the best from each and have a more powerful research team," says Teresa Fonseca Martin, a former particle physicist who spent seven years at ATLAS (she left this year to become a school teacher). "It is true that different cultures have different ways of working, but by paying a bit of attention, it is easy to learn about it and work with it."

Many of these opportunities involve learning softer skills, such as professional etiquette, leadership and management, communication and networking, and how research is conducted. These can be important for junior researchers who may be operating outside their home country for the first time and have had little contact with scientists from other countries. International consortia, says Fonseca Martin, also provide opportunities to develop a global network of colleagues and friends, as well as a chance to learn about the cultures of different countries.

**FIGHTING ANONYMITY**

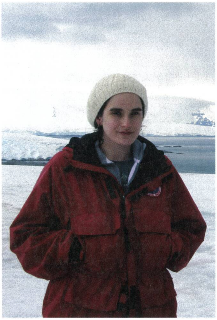
A big, prestigious team-science effort can not only boost a career but also sink one — or at the very least, waste the time of an early-career researcher. In particular, benefits can be offset by a numbing anonymity, especially for participants on the lowest rungs of the research ladder. The number of institutes and individual scientists involved turns large consortia into complex ecosystems that must be negotiated, whether researchers are trying to get credit for their lab work or attempting to stand out in a long list of names on a publication. Indeed, Ewan Birney, ENCODE's analysis coordinator at the European Bioinformatics Institute in Hinxton, UK, argues that the aims of individual participants in ENCODE and other big collaborations shift, from striving for excellent science that leads to publication and career success, to striving for maximum data output

## WHAT TO EXPECT

### *Look before you leap*

Early-career researchers in high-profile, international projects often struggle with how to stand out in a crowded field of graduate students and postdocs. Here are some tips to consider before — and after — joining a large project team.

- Seek advice about potential projects and principal investigators from knowledgeable advisers and researchers who have been associated with similar projects.
- Assess your personal and professional interest in the research, including whether the project will advance your career.
- Review the potential laboratory and research locations.
- Find out whether the principal investigator provides the mentorship and support you want.
- Seek opportunities for first-authorship on your own work within the project by carving out a special niche in the research.
- Look for chances to co-author publications with the principal investigator.
- Volunteer to take on administrative tasks for the project, such as writing papers, assisting in interviews and arranging meetings. This will help you to get your name recognized and acquire leadership skills.
- Accept opportunities to co-supervise PhD students with the principal investigator on discrete research projects within the collaboration.
- Try to discover something new in the research or to use a new technique that advances the project.
- Schedule regular meetings with the principal investigator to update him or her on any progress.
- Build good relationships with other early-career researchers on the project and set up meetings or web seminars to exchange information about their research.
- Look for chances to work at the main project site as well as at your home laboratory to raise your profile with senior investigators. **S.K.**


N. MURGAI

**Marine microbiologist Catherine Luria is part of a major ecology consortium in Antarctica.**

in the hope of contributing as much as possible to a community resource — usually a big data set (see page 49).

"Certainly there's an allure to a big project, but there's also a clear career risk of being lost in a very large crowd," says Julie Klein, who studies interdisciplinary teams at Wayne State University in Detroit, Michigan. "These are incredibly exciting and important projects, and they're seen as the future of science by some." They are also massive and unruly, she adds, in terms of the competition for attention. "It is often difficult to find one's place in a collaboration of 3,000 scientists," agrees Gonçalo. "At first it seems that every good idea you come up with has already been tried by someone else."

### STAND OUT IN THE CROWD

Nearly ten years after starting work on ENCODE, Jason Lieb, a biologist at the University of North Carolina at Chapel Hill and director of the Carolina Center for Genome Sciences at the university, says that standing out in a large team often means taking on extra work. He recommends that new members of the team improve their standing with the principal investigator by taking on extra roles, such as assisting in writing papers, hiring graduate students and scheduling group activities, and perhaps splitting their time between the large project and a smaller one in their home lab, with the aim of writing an independent paper with the principal investigator. Experienced postdocs say that developing leadership skills also helps a researcher to get noticed.

Another potential downside for the young scientist is the administrative effort required to operate these vast projects. For example, the scale of ATLAS, which includes about 3,000 physicists, has resulted in the development of an unhealthy, sluggish bureaucracy, says Fonseca Martin. These projects "don't necessarily get the best out of the people, and they sometimes make difficult the recognition of people's achievements and contributions," she adds, referring to assigning authorship and opportunities for promotion. Sometimes, says Fonseca Martin, a researcher's management abilities can become more important than their scientific ones.

Major collaborations often require much logistical effort, such as organizing meetings and conferences, notes Lieb. "People are tasked with certain jobs, and there's often a chance to take leadership positions in these jobs. If you're willing to try that, it's a good way to cut your teeth on a project." He adds that those who have taken on and performed effectively in such positions can demonstrate to their institute or university that they are team players who could, for example, make contributions to administrative tasks as tenured faculty members.

Along with taking on extra tasks, researchers can increase their profile by visiting and working in other labs involved in the collaboration. This helps them to build contacts and disseminate their research widely. "Projects that do better have postdocs or graduate students spend two or three months working in a lab at another site and then go back to their home institution," says Jonathon Cummings, who studies scientific collaboration at Duke University's business school in Durham, North Carolina.

But some researchers caution that graduate students and postdocs should be wary of becoming too closely associated with a single project, however glamorous, in case they become pigeonholed by peers and potential employers. "I worry that I'll be viewed as the 'person who works in Antarctica' and that will shape what I do later on," says Luria. "People are so interested in the place and fascinated by what we're doing, so it would be easy as a young scientist to have this experience become the defining quality of my work. I'm loving being in Antarctica and being a part of this project, but I'm trying hard to make sure it doesn't define me for the rest of my career."

Getting involved in a high-profile consortium can indeed be a headache, but it is often worth the effort, says Lieb. "People complain that these consortia are very clubby and difficult to get into," he says. "It's kind of true, but there's a reason why it's true. Once you've done it, you're more qualified to do it again. If you're able to get in early and demonstrate your skill at working on a project of this size, you're more likely to get another shot." ■

---

**Sarah Kellogg** *is a freelance writer in Washington DC.*

---

EUROPE

## Investment increases

Research and development (R&D) investment by European companies is on the rise, according to *The 2012 EU Survey on R&D Investment Business Trends*, a European Commission report released on 20 August. The survey of 1,000 large companies across all sectors predicts an average R&D boost of 4% a year until 2014. Chemical companies project an increase of 5.5%, and oil and gas producers 4.6%. "Employment costs are more than half of total R&D costs," says Alexander Tübke at the Institute for Prospective Technological Studies in Seville, Spain, a co-author of the report, "so an important share of R&D increases should translate into new employment." But, Tübke notes, any resulting researcher recruitment is likely to be in countries with lower labour costs, such as India and China.

---

EDUCATION

## Teachers lack resources

Full- and part-time teaching faculty members without tenure at US academic institutions face challenges that detract from their work and negatively affect their students, says a report released on 23 August by the New Faculty Majority Foundation in Akron, Ohio. A survey of 500 contingent faculty members found that they often don't know until days before a class begins that they are to teach it, and that most have no access to office or lab space, phones or computers. Such practices compromise students' educational experience, the report argues. Maria Maisto, executive director of the foundation, adds that uncertainty and lack of office space also hinder development of student–mentor relationships.

---

ENTREPRENEURSHIP

## Advice for protégés

To benefit from mentoring, fledgling entrepreneurs should be honest with their advisers about business issues such as cash flow; seek out mentors with similar values, personality or interests; and develop trust through frequent meetings, says a study based on a survey of almost 400 protégés (E. St-Jean *Int. J. Training Dev.* **16,** 200–216; 2012). Entrepreneurs who achieve good relationships with their mentors can build management knowledge and skills and improve their visions for their companies, says author Étienne St-Jean, who studies business management at the University of Quebec at Trois-Rivières in Canada.

**SPOTLIGHT ON WALLONIA, BELGIUM**

# 21st-century Marshall Plan boosts Belgium's capacity for innovation

## Six scientific clusters in the French-speaking Walloon region team academia with industry and raise the country's international R&D profile

> *"We would never have been able to grow as fast without the Marshall Plan, particularly since we started at a difficult time, just after the financial and economic crisis began."*
>
> ***Eric Halioua, CEO Promethera Biosciences***

**AN ODDLY-NAMED** Belgian economic initiative to boost the Walloon region's scientific edge is showing dividends. Though the Marshall Plan — originally the label for the US recovery program to rebuild postwar Europe — baffled some, few will argue that its effects have been positive.

The plan outlines spending about €4.3 billion over ten years to increase the French-speaking region's competitiveness and innovation. Phase one of the plan set the pace, with €1.6 billion spent from 2005 through 2009. The funds created five clusters: BioWin for life sciences, WagrALIM for agribusiness, MecaTech for mechanical engineering, Logistics in Wallonia for transport and logistics, and Skywin for aerospace. A sixth cluster for environmental science, Greenwin, has been set up in the €2.7 billion second phase.

Scientists have applauded the outcomes. The initiatives almost certainly mean more career oportunities, since the plan aims to strengthen research, spur economic growth, create new jobs and lift Belgium's R&D spending to the European Union (EU) target of 3% of gross domestic product (GDP) from about 2%.

NETHERLANDS
Flanders
Brussels
BELGIUM
GERMANY
Wallonia
FRANCE
LUX.

There is already evidence that the progam has been a success. The six-year-old life-sciences cluster BioWin claims to be one of Europe's leading groups in the field, with a growing international profile (see Inside View page 2). So far, BioWin has teamed up with three European Union (EU) clusters, some of its associated companies have sold their technology in the United States and China, and more international tie-ups are in the pipeline. "We are discussing possible partnerships between Walloon SMEs and companies in the Boston area of the US," says Sylvie Ponchaut, BioWin's scientific director.

Already Wallonia's three main universities — Université Libre de Bruxelles, Université Catholique de Louvain and Université de Liège, have produced several spin-offs under the BioWin umbrella, and are accompanied in the venture by other universities such as Mons and the Facultés Universitaires Notre-Dame de la Paix (FUNDP) in Namur.

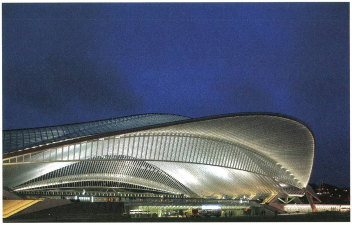
Industry-led, the clusters give Belgian companies access to generous public research funds for the first time. The business community, particularly small and medium enterprises (SMEs), had long lamented the lack of government grants. To obtain public funds, cluster projects must specify the number of jobs they will create, describe how they will attract additional investment, and include at least four partners — two companies and two research units from different universities.

Several examples point to the approach's potential: the Institute for Medical Immunology at Université Libre de Bruxelles (ULB-IMI), serves as a model for bringing industry and academia together; Promethera Biosciences demonstrates how locally grown stem cell technology can be developed; and the National Institute for Radioelements Environment & Lifescience Technology (IRE-ELIT), shows how a public utility can boost the radiopharmacology industry.

For Oberdan Leo, ULB-IMI director, the plan has altered attitudes about public-private financing. "In the past, industry asked university research labs to provide them with services," he says. "Now, projects are created through cooperation between industry and academia." In Belgium, almost all basic research is conducted within universities, which are funded based on the size of their student population.

The IMI was launched in 2004 as a private-public partnership with GSK Vaccines and the Walloon Region, One of Belgium's first PPPs in life sciences, the institute now boasts 800 square metres of lab space, including a biosafety level III unit, and technology platforms for flow cytometrey, cell culture, molecular biology and micro-surgery.

The immunology institute is not yet involved in a BioWin-funded project, but should form links imminently, Leo says. Plans are for immunologist Marcel Tuynder to create a spin-off company, BioCygnus, which would offer automated ex-vivo immune monitoring in whole blood, and a new diagnostic approach that



The futuristic Liege Guillemins TGV Station represents investment in Wallonia's R&D infrastructure.

uses molecular biology tools to evaluate new immunotherapies and measure vaccine efficacy.

"Although the timing is not ideal for raising capital to cover the first two years of running costs, we hope to create the company next January," adds Leo. Meanwhile, the BioCygnus lab has carried out assays for its product under contract with GSK Vaccines.

A more established BioWin company, Promethera Biosciences, demonstrates that cluster-created companies can thrive. The stem cell therapy firm was created in February 2009 to develop and market products to treat genetic liver disease. The first three patients were treated with Promethera HepaStem between March and June in a phase I/II clinical trial. If successful, the company says the injectable treatment could compensate for the shortage of donor livers available for transplants.

Another product, also based on the progenitor cell type Heterologous Human Adult Liver Progenitor Cell (HHALPC), is the Promethera HepaScreen. A non-therapeutic product, it uses the cells as a biotechnology tool to evaluate new chemical entities. Of the €12.2 million in loans and grants raised since the company was formed, €2.2 million came from research and investment grants provided by BioWin and the Walloon region, and €10 million in conditional loans from the region. The €12.2 million tops up the €22.3 million euros provided by international investors.

Promethera Biosciences CEO Eric Halioua says this varied support has allowed the company to thrive. "We now employ 46 people, including 41 full-timers, and we would never have been able to grow as fast without the Marshall Plan, particularly since we started at a difficult time, just after the financial and economic crisis began."

Apart from funding, an important element in the company's development is its neighbour GSK Vaccines, which provided four out of five of its key managers and several technicians and, Halioua says, helps keep Wallonia ahead of the pack for cell therapy in Europe.

The Marshall Plan has also given an appreciable boost to Promethera Biosciences' academic partner, the Université Catholique de Louvain (UCL). Although "it is difficult to put precise figures on it, there is no doubt it us helping researchers create their own companies more rapidly than they otherwise could," says Benôit Macq, one of the university's three vice-rectors and head of its information technology lab TELE.

"As far as I can see, BioWin and the five other thematic clusters are the biggest success of the plan," adds Macq, who is in charge of the service to society or applied research segment covering spin-off participations and creations, and patent follow-up. The other two are basic research and training. In 2011, the university's technology transfer office launched a second €45 million fund, Vives 2, which has taken a stake in Promethera Biosciences.

UCL has been responsible for 52 spinoff companies, of which eight have come from Macq's lab. They include the medical imaging company Telemis, which has received cash and technical assistance from BioWin. Much of the lab and spinoff work concerns information systems for medical support, adds Macq.

BioWin has financed 18 projects involving UCL, and has another three awaiting a decision. These involve 25 different research teams and have required the recruitment of 46 researchers and technicians, Macq explains.

While Promethera's success shows that Belgium has made the most of academic-industry alliances, a public sector initiative illuminates a novel way into the biotech industry. The subsidiary of a Walloon public utility foundation called IRE-ELIT launched two years ago to commercialize radiopharmacy. IRE-ELIT employs more than 20 people and has moved swiftly up the value-added chain to producing finished radioactive pharmaceuticals from straightforward bulk radioactive material, according to its business development manager Philippe van Put.

"We have BioWin to thank for our rapid progress innovation," he adds. The cluster has provided €5million for two projects. Of the total, €2.2 million has gone to RadioTarget to work on developing a treatment for liver cancer, and another €2.8 million has gone to GAGE to develop a diagnostic agent for neuroendocrine tumours. Both projects require a multidisciplinary input, which an SME on its own could not afford, explains van Put. BioWin has financed the recruitment of two full-time researchers and one technician, and has introduced partners. "If a link in the chain is missing, BioWin finds it," says van Put.

IRE-ELIT's ambitions extend well beyond Walloon's borders. It expects to export 95% of its output, mainly to Asia, the United States and throughout the EU. With help from the Chinese-Walloon innovation platform, BioWin and the Walloon export agency Awex, the firm is finalising a joint venture with a Shanghai-based company, which van Put prefers not to identify until the agreement is signed. The new venture will import radioelements and manufacture the finished products for direct distribution to hospitals, he adds. "The potential is huge, since nuclear medicine is not very developed in China."

The Marshall Plan also has a ripple effect. The National Fund for Scientific Research (F.R.S.-FNRS), which finances basic research, has no direct role in the plan's implementation, says director Véronique Halloin. But its associated Fund for Research Training in Industry and Agriculture (FRIA) awarded scholarships to 120 doctoral students working on industrial or agricultural themes under the plan's first phase. Meanwhile, the agency finances some 2,300 researchers, and in May 2011, an agreement was signed for the Walloon and FNRS national contact points or portals to cooperate by focusing on their complementary strengths to help scientists obtain European Union grants.

Even though government budget constraints may affect R&D investments, arrangements under the Marshall Plan have laid foundations for the Walloon scientific community to make the best of a dynamic pool of knowledge and a beneficial collaborative spirit. ■

*Nature editorial staff have no responsibility for content*

### GIGA illustrates Marshall Plan reach

Wallonia's scientific clusters did not first appear with the Marshall Plan. One biomedical grouping, GIGA, which is hosted by the University of Liège and houses Wallonia Biotech Coaching, was launched in 2003. (See Inside View page 4).

The life sciences institute WELBIO is financing three two-year GIGA projects, GIGA teams are involved in nine of the 25 BioWin projects, and the government has put up €10 million of the €25 million invested in the NEOANGIO cancer therapeutics 'programme for excellence'. "This is not an exhaustive list, but gives an idea of the Plan's reach," says Catherine Sadzot, the GIGA general coordinator. Beyond that, GIGA's technological platforms and other infrastructure have been mostly co-financed by European Regional Development Fund (ERDF) and the Walloon Region.

BioWin
THE HEALTH CLUSTER OF WALLONIA

# :insideview
## profile feature

**Professor Jean-Luc Balligand,** *professor of medicine and pharmacology, head of the pole of pharmacology and therapeutics, at the Université Catholique de Louvain and BioWin Vice-President*

*Standing out in the jungle of European innovation clusters is no mean feat. Thanks to its regional strategy, which builds on its academic and industrial areas of excellence, BioWin has just done that. Based in the Wallonia region of Belgium, the health cluster now boasts the first success stories from its public-financing support programme, designed to create public-private partnerships between academia and the private sector. With members generating an annual turnover of €3.4 billion and employing approximately 12,500 people, BioWin is one of the top European competitiveness clusters to date, rapidly acquiring worldwide recognition. BioWin differentiates from other biotech clusters in Europe through its integrated support programme, which combines R&D project financing with a proactive approach to training and skills development. The cluster also offers initiatives designed to meet the needs of industry partners. In addition, BioWin deliberately focuses on providing technology platforms to support select areas of research excellence, including cell and proton therapies.*

*After six years of activity, the cluster is ready to pursue its international development strategy and firmly put the Wallonia region on the international biotech map. Some projects have already come to fruition, with BioWin-associated companies selling technology developed in Wallonia as far afield as the U.S. and China. BioWin vice-president Jean-Luc Balligand explains the strategies behind the cluster's success.*

### Q. What makes BioWin so attractive to its members?

Supported by the regional government of Wallonia, BioWin has set up unique policies that favour interactions between academics and small and medium enterprises (SMEs) in the health biotech and/or MedTech sectors as well as global companies flourishing in the region such as GSK Vaccines, UCB, IBA and Baxter. Instead of working independently, these entities now join forces to develop innovative research typically at the core of SME's business. Academic research laboratories typically get 100 percent of their expenses covered by regional funding, while industry partners get between 50 and 80 percent of their costs subsidised and are expected to cover the rest through private contributions. BioWin currently has 510 members, including around 100 innovative companies, among which several spin-offs have emerged from three university academies centered on the three main universities (Université Libre de Bruxelles, Université Catholique de Louvain and Université de Liège) and accredited research centres.

The support provided to translational research is also a very attractive factor. For example, BioWin initiated two shared technology platforms: one in proton therapy (still in development) and one in cell therapies. The latter is provided through the Charleroi-based company MaSTherCell, which manufactures cell-therapy products and technologies for member SMEs. As a result, the economic value added by BioWin member companies has more than doubled since 2006, a sign of vitality that also attracts new members.

### Q. How do you ensure that the various partners can hire adequately trained scientists?

BioWin has developed 14 tailor-made training programmes to renew and update the skills of scientists working in biotechnology. This allows qualified professionals to stay current in their field, learn about the latest research and experience work in an industry setting. As a result, 75 percent of in-company trainees found a job in the industry within three months and 100 percent of participants to the CRO-training ClINET also got a job in a Contract Research Organisation. Moreover, to ensure that skills development proposed by different training operators meet the demand of the biotech health sector, BioWin has developed a training framework program, the BioPhare project (BioPharmaceutical Excellence) to structure training programs and match the evolving needs of its members.

### Q. What strategies have enabled BioWin to punch above its weight internationally?

BioWin has a research strategy based on expansion through international collaborations with other biotech clusters in Europe. We are active in three EU inter-clustering programmes (TERM, ABC Europe and Nanofar), which involve weekly interactions with 30 other European health clusters. BioWin is also acting as vice-chair of the Council of European BioRegions (CEBR), which comprises 40 health clusters and scientific parks. WAL-Dx, the *in vitro* diagnostics (IVD) branch of BioWin, founded the European Diagnostic Clusters Alliance (EDCA). This placed Wallonia at the geographical heart of a trans-European collaboration in its specific area of excellence.

In addition, BioWin is currently represented in Boston and Shanghai with dedicated teams that forge partnerships with several international organisations and institutions outside Europe, which helped boost our international profile. These partners include the Texas A&M University system in Houston, as well as the Shanghai Biopharma Industry Association (SBIA) and the Juke Biotech Park, also in Shanghai. One of our members, IRE-ELiT — a spin-off from the Institute for Radio-Elements in Fleurus — is now involved with a radiopharmacy project in China that uses Rhenium Re188 produced from Tungsten W188 generators built in Wallonia.

### Q. What results have you seen from BioWin's first six years of activity – and what are your next steps?

Further concrete success stories include the development of new formulations of inhaled antibiotics with large applications to chronic respiratory diseases and another for local treatment of precancerous lesions of the cervix, for which a phase II trial is being finalized. Meanwhile, a UK-based company has finalized a deal to commercialize holographic microscopes built in Wallonia for real-time monitoring of bioreactors . Overall, this has secured the recruitment of more than 200 researchers with 1,250 more jobs to be created.

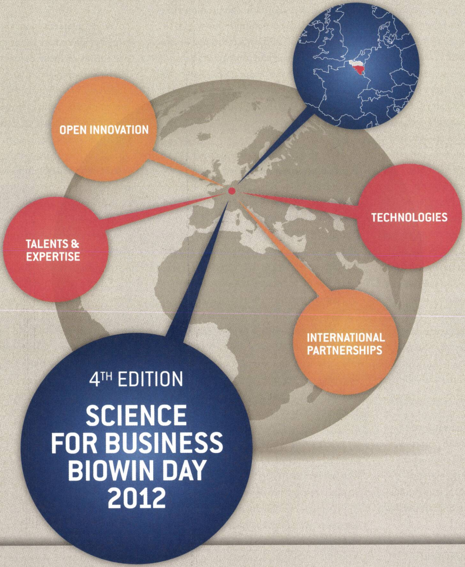
In the area of cell therapies, for example, applications range from cardiology to bone and liver diseases, developed by start-up companies such as Cardio 3 BioSciences, Bone Therapeutics and Promethera BioSciences, respectively.

The next step is to attract foreign industry partners and ensure we retain the value created through such partnerships in Wallonia.

ADVERTISER RETAINS SOLE RESPONSIBILITY FOR CONTENT

# BIOWIN,

# WHERE SCIENCE MEETS BUSINESS



OPEN INNOVATION

TECHNOLOGIES

TALENTS & EXPERTISE

INTERNATIONAL PARTNERSHIPS

4TH EDITION

**SCIENCE FOR BUSINESS BIOWIN DAY 2012**

THEME:

## INNOVATION FOR THE ECONOMY AND FOR THE PATIENTS

## NOVEMBER 28TH

**AULA MAGNA > LOUVAIN-LA-NEUVE > BELGIUM**

CONFERENCES - WORKSHOPS - EXHIBITION - JOB CORNER
POSTER SESSION - SPECIAL EVENING WALKING DINNER

BE PART OF IT! **WWW.BIOWINDAY2012.ORG**

BioWin
THE HEALTH CLUSTER OF WALLONIA

# :insideview
## profile feature

**Jacques Piette,** *Research Director, GIGA-Research, University of Liège*

*The desire to build stronger links between scientists and clinicians, and to translate basic biology research into clinical outcomes, led to the creation of The Groupe Interdisciplinaire de GénoProtéomique Appliquée (GIGA) at the University of Liége. Before GIGA's creation in 2003, biomedical research was spread across the university's campus. Since then, clinical translational medical research has been concentrated within the Faculty of Medicine and the Centre Hospitalier Universitaire de Liège (CHU), and the GIGA has grown to comprise 550 researchers from four faculties (Medicine, Sciences, Applied Science and Veterinary Medicine) structured around seven thematic research units. It prioritises technological platforms, offers business support and facilities for biotech companies, including intellectual property support and patent advice, and nurtures spin-off companies and start-ups. It also provides training in bioscience techniques to job hunters, company staff and college students. Jacques Piette, Research Director, explains how and why the GIGA offers an exceptional environment in which to conduct state-of-the-art academic biological research and launch a successful career or business.*

**Q. What is the GIGA's research focus?**
We are focused on genomics and proteomics, across the following areas: cancer; genetics; development, stem cells and regenerative medicine; infection, immunity and inflammation; neurosciences; signal transduction; and, systems biology and chemical biology.

**Q. What makes the GIGA unique?**
GIGA offers a research center with technical platforms and equipment managed by experts; an intellectual property (IP) centre; a business infrastructure that supports researchers in creating spin-off companies based on aspects of their research; and, a training center in Biotechnology.

The research centre enjoys close proximity to the CHU, and this aspect is unique in Wallonia. The location enables close collaboration between researchers in biomolecular sciences and clinicians, and increases the number of projects involving the two camps.

The GIGA's 'open innovation' model includes open labs, and facilitates the transfer of ideas and skills between disciplines, as well as between researchers and clinicians. This promotes free thinking and idea generation. Our on-site business support and infrastructure encourages spin-off companies and startups. Staff members from our Intellectual Property centre visit the labs to answer queries and help the researchers navigate their way through the often daunting patent process.

**Q. Why should talented postdocs set their sights on joining the GIGA?**
We provide high-end laboratories and equipment, and the opportunity to engage in exciting collaborative research. This, in turn increases the likelihood of publication in top journals. Researchers in basic biology can conduct translational research, collaborate with clinicians and see work translated into clinical outcomes and patient benefits. The technology platforms comprise state-of-the-art equipment run by experts and are available to all researchers. Aside from the science opportunities, the GIGA offers a welcoming, international environment, with language support, social support, career development, as well as further business and teaching opportunities.

**Q. What skills are you looking for?**
Currently we are particularly in need of people with bioinformatics and data handling expertise. So much research involves interpretation of complex data and skills perhaps not traditionally associated with biology research. GIGA also wants those researchers keen to work in a multidisciplinary environment and those with an appetite for translational research — the cross-over of molecular and cellular biology with health and medicine.

**Q. What exciting and unusual collaborations have you seen between contrasting disciplines at the GIGA?**
Collaboration between scientists in different disciplines is to be expected today. Much more exciting to me is the unusual use of or creative approach to problem solving, evidenced when people use skills/models/techniques in unexpected situations, for example combining random mutagenesis, whole genome sequencing with the zebrafish model to better understand organ development in humans.

**Q. What projects have the highest profile right now? What are the expected outcomes of these projects?**
We have many high profile research projects underway. Among them, I would highlight projects led by a young neurobiologist carrying out exciting work to elucidate the molecular regulation of neurogenesis in the brain. This group is actively collaborating with other GIGA researchers to better understand the role of a protein complex that seems to be important both in brain neurogenesis and in cancer development.

Another team working on asthma has identified novel mucosal innate immune cell populations. This group has also recently published a work in collaboration with a Japanese team from Osaka University on the importance of the DNA released by dying cells in mediating the adjuvant activity of aluminium salts, which could lead to new vaccination strategies.

Others are integrating genetics and functional genomics to identify causative genes and variants controlling inherited predisposition to inflammatory bowel disease.

**Q. What next?**
A new cardiovascular sciences unit will be up and running within the next two years and will conduct mainly clinical research. The aim is to gather clinicians interested in research to deliver a truly bench-to-bedside approach.

**Q. What is your vision of the GIGA's future?**
I aim to grow the number of projects that translate basic (molecular and cellular) biological research into clinical outcomes, increase the number of projects that involve international collaborations and publish a higher proportion of our research in journals with high impact factors. The centre's long-term future will be intertwined with the fortunes of science funding in Belgium and in Europe, and reliant on the GIGA's existing and future success. Our ultimate vision is for the GIGA to grow in stature, not necessarily in size, and to build its reputation overseas, so that it becomes a prestigious launch pad to a career in top-level international science.

ADVERTISER RETAINS SOLE RESPONSIBILITY FOR CONTENT

# The GIGA, more than a research center

The GIGA is a unique organizational structure in Belgium located within the University of Liège. Its interdisciplinary, its location, in the very heart of the University hospital allowing a constant interaction with the clinicians are undeniably its major assets. Indeed, the GIGA was created to gather researchers from different disciplines in one location whose architecture was designed to optimize interactions while integrating an area dedicated to training for workers in the biomedical sector and opening doors to biotech companies.

## The academic research center

More than 400 researchers (20% foreigners) working in close collaboration with clinicians on clinically-oriented research topics :

**7 Thematic Research Units**

- **Genetics**
- **Cancer**
- **Neurosciences**
- **Signal Transduction**
- **Infection, Immunity & Inflammation**
- **Development, Stem cells & Regenerative medicine**
- **Systems biology & chemical Biology.**

An 8th research unit focused on Cardiovascular Sciences will be created in September 2012.

## Translational research : from patients to benches and from benches to patients

The ideal location within the academic hospital eases the collaboration with the clinicians and leads to translational research programs.

## The technology platforms

GIGA-Technology platforms gather up-to-date technologies managed by experts who are fully dedicated to the Platforms. All services and equipments are open to academic researchers as well as to the private sector, contributing to speeding up the biomedical research.

- **Genotranscriptomics**
- **Proteomics**
- **Imaging and flow cytometry**
- **Bioinformatics**
- **Mouse facility and transgenics**
- **Zebrafish facility and transgenics**
- **Immunohistology**
- **Viral vectors**
- **Interactomics**
- **BioBank**

## The business facilities

The GIGA has 3 Business Facilities areas aiming to offer a place of development to biotech companies (spin-offs, start-ups or subsidiaries of established companies). In addition, a «labhotel» located within the research center, equipped with basic equipment and allowing access to the whole infrastructure has recently opened.

## The training center

Embedded within the research center, the training center offers training sessions for job seekers, high school students or workers. Customized training can also be designed to answer the specific needs of a company.

## Any opportunities to join us ?

As an interdisciplinary center, we are frequently searching both for research scientists (biochemist, biologist, pharmacologist), bioinformaticians, engineers, or technicians interested to work in an international, very interactive and stimulating environment.

We are actually searching for 5 PhD to join the groups working on Cancer, Animal Genomics and Neurosciences.
A position for a postdoctoral medical bioinformatician for an integration of Next Generation Sequencing data into clinical settings is also available.

Job offers for PhD students, Postdocs or technicians are frequently posted on our web site : **www.giga.ulg.ac.be/jobs**

## How to learn more about GIGA ?

**www.giga.ulg.ac.be**
**facebook.com/GIGAresearch**
**twitter.com/GIGA_ULg**

**GIGA**
Interdisciplinary Cluster in Applied Genoproteomics
University of Liège
Avenue de l'Hôpital 1 – B34
4000 Liège – Belgium
Tel +32 4 366 98 32
info.giga@ulg.ac.be

welbio

**Walloon Excellence in Life Sciences and Biotechnology**

WELBIO is an inter-university research institute based in Wallonia, Belgium. WELBIO aims at promoting scientific excellence in **fundamental life sciences research** and translating scientific achievements into medical, pharmaceutical and veterinary **biotechnology applications.**

Founded in June 2009 as a non-profit association, WELBIO is supporting **15 research programmes** in the three Walloon academies through its first call for projects in 2010. Additional programmes will be funded through further calls for projects.

The institute regularly provides outstanding opportunities for **postdoctoral candidates** to join our research programmes and take part in innovative research in life sciences.

More information on **http://welbio.org**

W227957R

---

Université de Liège

**Academic position in Systems Biology and Bioinformatics**
**Department of Life Science – Faculty of Sciences**
**University of Liège**

The University of Liège (Belgium) seeks applications for an academic position in the area of Systems Biology and Bioinformatics in the Department of Life Sciences, Faculty of Sciences (www.facsc.ulg.ac.be). Applicants must possess a PhD (preferably in the biological sciences), a distinguished record of scientific research and extramural funding, administrative and leadership skills, and a strong commitment to education and academic excellence. Applicants should be able to teach in English but are also expected to teach in French, possibly after a transition period. The candidate will work in the department of Life Sciences and will be expected to foster new areas of research while supporting and enhancing current research strengths.

Nominations and inquiries may be directed to **Mme F. MOTTE, Faculty's Administrative Manager - tél. : +32 4-366 36 52 – Francoise.Motte@ulg.ac.be**. Candidates should submit, by registered letter, their application to **Professor Bernard Rentier, Rector, University of Liège, Place du 20 Août 7, B4000 Liège, Belgium**. The application should be sent before **October 1st 2012** and must include a letter of interest, a curriculum vitae, a complete list of publications and a research project of five to ten pages.

W226768R

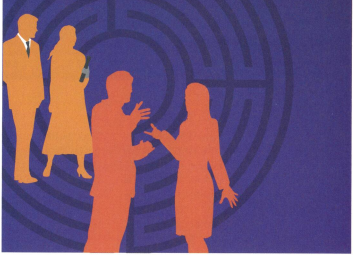
---

naturejobs

**Attention Employers!**

Stay up to date with deadlines and upcoming ***Naturejobs*** features:

Sign up for **FREE** employer e-alerts:
**naturejobs.com/ealerts**



---

**We would like to use this opportunity to thank all Universities, Clusters and Government organisations for their support and participation in the 'Spotlight on Wallonia'.**

**The 'Spotlight on Wallonia' will be available on naturejobs.com**

**Please visit our website for career information and advice.**

**For ongoing recruitment advertising enquiries, please contact us at the details below.**

**Thank you,**

**Julie Aramburu**
Senior European Advertising Executive
Belgium | France | Switzerland | Luxembourg
T: +44 (0) 20 7843 4994 | E: j.aramburu@nature.com

naturejobs

nature publishing group npg

Web Search Images Maps News Books Translate ▼

Search help | Search settings | Sign in

UNIVERSITY Search

UCL - Université catholique de Louvain | Search

Advanced Search

Announces

### Research performance: silver medal for Belgium

UCL researchers rank high thanks to the Belgian «Interuniversity Attraction Poles» (IAP) Programme. A study describes why Belgian researchers rank among the world's highest thanks to the IAP programme.

http://www.uclouvain.be\michel.gevers\BelgianResPerf.Pdf

http: www.belspo.be\IAP

### Examples of UCL's research structures that improve people's life

#### WELCOME: Wallonia Electronics and COmmunications MEasurements

UCL's WELCOME facility is a new characterisation platform that supports innovation in electronics and telecommunications fields.

http://www.uclouvain.be\welcome

#### WINFAB, UCL unveils its new micro-nanofabrication lab

WINFAB is UCL's micro-nanofabrication platform which provides researchers and students with a cleanroom environment and high-tech equipment.

http://www.uclouvain.be\en-winfab

#### Hoover Chair celebrates 20 years

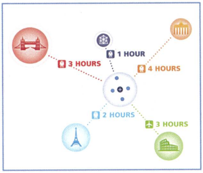
In the last 20 years, UCL's Hoover Chair has become a major player in ethical issues related to the economic and social sphere.

http://www.uclouvain.be\en-chaire-hoover

### Images for «UCL» - Report images

UCL
Université catholique de Louvain

Netherlands
United Kingdom
Belgium
France
London
Paris
C
B
E
F
G
H



Announces

### EURAXESS encourages researcher mobility

This UCL-European Commission initiative gives researchers the opportunity to advance their career by facilitating mobility.

http://www.uclouvain.be\en-euraxess

### UCL contributes to 6 million euro fund to host foreign post-docs

UCL on behalf of Académie universitaire Louvain has joined forces with the European Commission in the first large-scale funding of foreign post-docs.

http://ec.europa.eu/euraxess/index.cfm/jobs/fgDetails/33321

### Louvain International Desk (LID)

The LID helps foreign researchers and their families at each step of their international mobility. This personalised service aims to provide up-to-date information as well as individualised assistance in all areas, whether administrative or legal, or concerning everyday life in the new host community.

http://www.uclouvain.be\en-lid

- F Welcome & WinFab
  10 min. walking
- G Chaire Hoover
  10 min. walking
- H TECLIM
  10 min. walking
- E de Duve Insitute
  30 min. by bus
- B Paris
  2 hours by train
- C London
  3 hours by train

UCL
Université catholique de Louvain

## There's one solution for all your research and personal needs: UCL

Researchers head toward the heart of Europe to benefit from a privileged environment. Renowned for its continuous pursuit of quality, innovation and scientific research, UCL has earned a reputation as a world-class institution that pampers researchers and caters to all their needs. Share knowledge and understanding in an inviting and inclusive environment with the best minds of Belgium and beyond.

**www.uclouvain.be**

W228139R

# :insideview

## profile feature

Contact Singapore
contactsingapore.sg

**Hilmi Volkan Demir,** *Singapore National Research Foundation Fellow, Nanyang Technological University of Singapore*

*Young, independent and funded—that's the status Hilmi Volkan Demir achieved when he won Singapore's National Research Foundation (NRF) Fellowship. Unlike junior professors, Demir did not have to worry about securing additional grants immediately upon arriving at the Nanyang Technological University of Singapore (NTU). Since NRF Fellowships provide up to S$3 million (about US$2.4 million) over five years, recipients can support a team of postdocs and PhD students, as well as research equipment, expenses and travel. The Fellowship scheme is open to young scientists from around the world. Besides the freedom from securing start-up funding, NRF Fellows can choose which Singapore university or research organization they will work at and have independence to pursue their own projects—as long as they exhibit a strong likelihood of a scientific breakthrough. Although there is no limit on the number of Fellowships available each year, there are typically about ten awarded. Since the launching of the NRF Fellowship program, there are now close to fifty Fellows in Singapore. Demir tells* Nature *how the NRF Fellowship jumpstarted his career in Singapore.*

### Q. Why did you apply for the NRF Fellowship?

I applied for the NRF programs because NTU contacted me and strongly encouraged and supported me to take part in the NRF Fellowship and Competitive Research Programs. Here, NTU provides a strong enabling environment essential to top-notch research, with world-class facilities including cleanroom infrastructure, powerful research support systems, and top-rated scientists and colleagues. Also, NTU and Singapore in general, strongly encourage close collaboration across the university and around the globe, which allows me to work closely with my native country Turkey, as well as Europe and Northern America. Both NRF and NTU give us the full freedom to pursue our own scientific research programs. Here there is a lot of encouragement for scientific excellence, innovation and entrepreneurship. These were personally very appealing to me.

### Q. Why is the NRF Fellowship especially attractive to young researchers?

Both the NRF and Singapore want to see young scientists succeed and develop their own research program. This requires financial resources, quality students and postdocs, space and research infrastructure. In Singapore, you can get the funding you need, the people you need, the space you need, and the lab environment you need.

All these elements together form an ecosystem. Many young scientists consider themselves fortunate to start up with one or two of these elements. In Singapore, NRF, together with the hosting universities, basically enables this ecosystem. Here it's a matter of efficiency. Singapore provides all the parts, because otherwise, for every missing component, you end up spending your own time and your own resources.

### Q. What has winning the NRF Fellowship done for your career?

NRF programs have allowed me to think about technological impact and potential innovation while focusing on the excellence of science. With NRF, we established a pioneering Center of Excellence for Semiconductor Lighting and Displays, called *Luminous!* with LED MOCVD growth, fabrication and testing at NTU Singapore. This center combines the expertise and scientific efforts of more than 30 researchers dedicated to high efficiency, high quality light sources and systems. We have got our labs up and running in under a year, which is extraordinary considering the scale of our operation and infrastructure here. Last year NRF funded Luminous! at about S$10 million (8 million USD) over five years through another NRF program called NRF Competitive Research Program (CRP). This shows the strengths of collaborations within the Singapore scientific community, which I think should be especially appealing to young scientists. In Singapore, NRF Fellowscan find quality collaborations and secure further funding beyond their fellowship period with various funding programs.

NRF programs have surely helped us to challenge ourselves and push our boundaries, and boosted our creativity in many ways. Here we are strongly encouraged to propose new ideas and make new proof-of-concept demonstrations that offer the potential to become game-changers. The expectation is very high. That drives us to work more together and seamlessly with rest of the world. In NRF programs, our team has been very productive at Luminous!, and we have been making important contributions to LED and energy research in the region and worldwide including energy-saving photometric quality LEDs, excitonic LEDs, remote nanophosphors of quantum dot sheets, and polarized emission of plasmonic quantum dot composites.

### Q. How else does the NRF help young scientists' careers?

NRF not only grants high profile research programs, but also helps you monitor your progress and receive periodic feedback, which is very encouraging. NRF has been operating very openly and is doing a great job at supporting us. The quality and impact of our work are evaluated by an international scientific advisory board composed of independent, very established and prominent individuals both from academy and industry, with whom we get to meet at various formal and informal gatherings and events and talk with individually. Various performance indicators are also used, some of which include the amount and potential of commercialization and the technological impact and enabling character of our research work. NRF further aims to proactively integrate the entrepreneurship landscape including both private and state investment with the granted research programs, which I think is very unique, and may be indeed quite appealing to young scientists interested in pursuing entrepreneurship activities complementing their scientific efforts.

Start your search for career opportunities in Singapore:
**www.contactsingapore.sg/jobs**

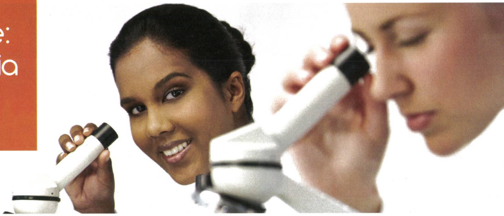
**Contact Singapore engages overseas Singaporeans and global talent to work, invest and live in Singapore.**
For more information on working and living in Singapore, please visit **www.contactsingapore.sg**
You can also email our officers in London: **london@contactsingapore.sg**
For more information about the NRF Fellowships, please visit **www.nrf.gov.sg**
or email Ms. WONG Wan Hui: **wong_wan_hui@nrf.gov.sg**

ADVERTISER RETAINS SOLE RESPONSIBILITY FOR CONTENT

# Careers@Singapore: Research & Academia

17 – 21 September 2012

United Kingdom



## Come to Singapore! Exciting career opportunities in Research & Academia await you

Singapore is brimming with opportunities for bright and ambitious research talent and academics across a broad base of research fields at all levels from PhD Scholars, Research Fellows and Clinician Scientists to Translational Researchers, Principal Investigators and Professors (Assistant/ Associate/ Full).

Contact Singapore and Singapore's top research institutes and world class universities will be in the UK for a series of recruitment events:

A*STAR Institute of Chemical and Engineering Sciences

NANYANG TECHNOLOGICAL UNIVERSITY

SUTD SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN Established in collaboration with MIT

**Sign-up for *Careers@Singapore: Research & Academia* (17-21 Sep 2012) to find out more about opportunities in Singapore: www.contactsingapore.sg/CareersSG/Research_Academia/Sep2012**

| Date | Time | Venue |
|---|---|---|
| Mon, 17 Sep | 1230 - 1400 | University of Warwick, Coventry |
| Tue, 18 Sep | 1300 - 1430 | University of Cambridge, Cambridge |
| Wed, 19 Sep | 1300 - 1430 | University of Oxford, Oxford |
| Wed, 19 Sep | 1830 - 2000 | Institute of Physics, London |
| Thu, 20 Sep | 0900 - 1700 | Naturejobs Career Expo 2012, London |
| Fri, 21 Sep | 1800 - 2000 | Holiday Inn Kensington Forum, London |

Not able to meet us in person? Join us at our upcoming **Contact Singapore Virtual Career Fair for Research and Academia on 12 and 13 October 2012**. Visit **www.contactsingapore.sg/Research/VCF/Oct2012** to find out more about the event.

**Contact Singapore**

Contact Singapore, an alliance of the Singapore Economic Development Board and Ministry of Manpower, engages overseas Singaporeans and global talent to work, invest and live in Singapore. We actively link Singapore-based employers with professionals to support the growth of our key industries. Contact Singapore works with investors to realise their business and investment interests in Singapore.

To start your search for career opportunities in Singapore, visit Contact Singapore's job portal at **www.contactsingapore.sg/jobs**

For more information on working and living in Singapore, please **visit www.contactsingapore.sg**

You may also email our officers in London: **london@contactsingapore.sg**

Contact Singapore

contactsingapore.sg

naturejobs
CAREER
EXPO
LONDON 2012

# 20 SEPTEMBER 2012

## BUSINESS DESIGN CENTRE, LONDON

## PREVIEW FEATURE

Dear *Naturejobs* reader,

This feature presents the 2012 *Naturejobs* Career Expo, brought to you by *Naturejobs,* the careers and recruitment section of Nature Publishing Group. The Expo will take place on Thursday 20th September at the Business Design Centre, Islington, London.

Since 2009, the Expo has been the UK's largest annual career fair aimed at all those in the physical sciences, life sciences and medicine. This year's event – the sixth in the series – is the largest yet. On 20th September, we welcome over 50 companies and organisations to the exhibition hall. These exhibitors will be presenting a very broad range of career opportunities, both in the UK and abroad. All the exhibitors are actively seeking qualified scientists to fill a breadth of positions and are eager to talk to you about your skills.

During the day, five organisations will be running in-depth workshops to tell you more about their aims, scope and job opportunities.

Alongside the exhibition and workshops, the conference programme offers a broad range of topics, presented by speakers from the public and private sectors, and from academia and industry. We are delighted to welcome as our keynote speaker Jorge Cham, Creator/Writer, PHD Comics. Following the keynote are three concurrent tracks— with over 15 sessions to choose from, with topics to include careers in academia and industry, managing your online presence, transferable skills and a lively session on networking.

We would like to thank all the individuals and organisations who have helped in developing the conference programme and format of the Expo and all our academic ambassadors. We also thank our Event Partner, our Gold and Silver sponsors, and all the other exhibitors for their support.

The Expo is designed for you, the scientist, to explore, discuss, and seek out the opportunities that will take you on the next step of your career. Attending the Expo is free, and there is a small fee to attend the conference. To register, please go to **www.nature.com/natureconferences/njce2012**

We look forward to seeing you on the day.

*Naturejobs*

**Attendance is free; register at www.nature.com/natureconferences/njce2012**

Gold Sponsors:

QBRI
Qatar Biomedical Research Institute

crp SANTÉ
CENTRE DE RECHERCHE PUBLIC

Research in Germany
Land of Ideas

Silver Sponsors:

VIB

Otsuka
Otsuka-people creating new products for better health worldwide

BASF
The Chemical Company

WORK IN DENMARK

MERCK

Register at www.nature.com/natureconferences/njce2012

Contact Singapore
contactsingapore.sg

naturejobs

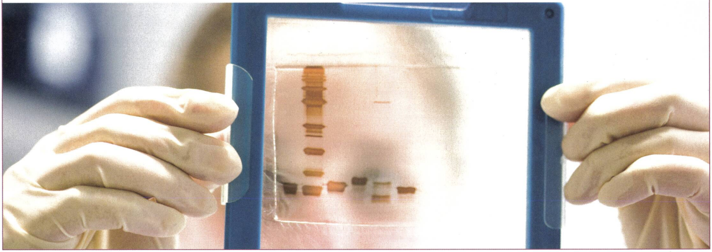
nature publishing group npg

W227302R

www.crp-sante.lu

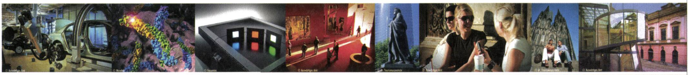
crp SANTÉ
CENTRE DE RECHERCHE PUBLIC

# CRP-SANTÉ: ALWAYS SEARCHING FOR MORE

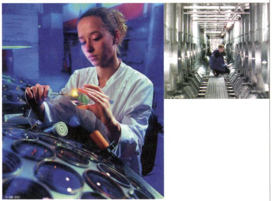


---

# Let's talk about your research career in Germany!



Register now for our **Meet and Greet** with German graduate schools and universities on **21 September** in London. Schedule **one-on-one meetings** with German representatives and learn about vacant PhD and postdoctoral positions.



Meet and Greet
21 September 2012
Register now!
www.research-in-germany.de/london2012

AN INITIATIVE OF THE
Federal Ministry
of Education
and Research

**DAAD** Deutscher Akademischer Austausch Dienst
German Academic Exchange Service

Research in
Germany
Land of Ideas
www.research-in-germany.de

W227332R

naturejobs
CAREER
EXPO
LONDON 2012

# 20 SEPTEMBER 2012

## BUSINESS DESIGN CENTRE, LONDON

## EXHIBITORS

**academics.com**
www.academics.com

**Baden-Württemberg International**
www.bw-career.de

**BASF – The Chemical Company**
www.basf.com

**BGI**
hr.genomics.cn/hr_en

**Bioscience Careers Group**
www.societyofbiology.org

**BIOSS Centre for Biological Signalling Studies**
www.bioss.uni-freiburg.de

**bizkaia:xede**
www.bizkaiaxede.org

**Cancer Research UK**
www.cancerresearchuk.org

**Centre de Recherche Public de la Santé (CRP-Santé)**
www.crp-sante.lu

**Contact Singapore**
www.contactsingapore.sg

**EMBL**
www.embl.org

**European Commission; EURAXESS-Researchers in Motion**
ec.europa.eu/euraxess

**European Commission; European Research Council (ERC)**
erc.europa.eu

**European Commission; Marie Curie Actions**
ec.europa.eu/mariecurieactions

**French Embassy**
www.ambascience.co.uk

**Fujifilm Diosynth Biotechnologies**
www.fujifilmdiosynth.com

**Georg-August-Universität Göttingen**
www.uni-goettingen.de

**German Academic Exchange Service (DAAD)**
www.daad.org.uk

**German Cancer Research Center (DKFZ)**
www.dkfz.de

**German Research Foundation (DFG)**
www.dfg.de

**Graduate School of Excellence "MAterials science IN mainZ" (MAINZ)**
www.mainz.uni-mainz.de

**Hochschule Mannheim University of Applied Sciences**
www.hs-mannheim.de

**IK4 Research Alliance**
www.ik4.es

**The Institute of Chemical & Engineering Sciences (ICES)**
www.ices.a-star.edu.sg

**International Giessen Graduate School for the Life Sciences**
www.uni-giessen.de

**International Graduate School in Molecular Medicine Ulm**
www.uni-ulm.de

**International Max Planck Research School for Neural Circuits**
www.imprs.brain.mpg.de

**Johnson & Johnson**
http://careers.jnj.com

**Karlsruhe School of Optics & Photonics**
www.ksop.de

**Karolinska Institutet**
www.ki.se/careerservice

**Konstanz Research School Chemical Biology**
www.chembiol.uni-konstanz.de

**Macmillan Science Communication**
www.mscediting.com

**Max Planck Society**
www.mpg.de

**The Merck Group**
www.merckgroup.com

**Nanyang Technological University (NTU)**
www.ntu.edu.sg

**Nature.com Open Innovation Pavilion**
www.nature.com/openinnovation

**Naturejobs**
naturejobs.com

**Nature Publishing Group**
nature.com

**Otsuka Pharma**
www.otsuka.co.jp

**Philipps-Universität Marburg**
www.uni-marburg.de

**Qatar Biomedical Research Institute**
www.qf.org.qa

**Research in Germany - Land of Ideas**
www.research-in-germany.de

**Singapore University of Technology and Design**
www.sutd.edu.sg

**Switzerland**
www.swissuniversity.ch

**Talent Scotland**
www.talentscotland.com

**Teach First**
www.teachfirst.org.uk

**Technical University of Denmark (DTU)**
www.dtu.dk

**Technische Universität München**
www.tum.de

**UBC Envision**
www.ubc-envisiongroup.com

**University of Copenhagen**
www.ku.dk

**University of Münster**
www.uni-muenster.de

**VIB**
www.vib.be

**Vitae**
www.vitae.ac.uk

**Workindenmark**
www.workindenmark.dk

**Attendance for the exhibition is free; register at www.nature.com/natureconferences/njce2012**

Gold Sponsors:

QBRI
crp SANTÉ CENTRE DE RECHERCHE PUBLIC
Research in Germany Land of Ideas

Silver Sponsors:

VIB
Otsuka
BASF The Chemical Company
WORK IN DENMARK
MERCK

Register at www.nature.com/natureconferences/njce2012

Contact Singapore
contactsingapore.sg

naturejobs

nature publishing group npg

QBRI
معهد قطر لبحوث الطب الحيوي
Qatar Biomedical Research Institute

Member of Qatar Foundation عضو في مؤسسة قطر

# EXCELLENT JOB OPPORTUNITIES IN QATAR

Qatar Biomedical Research Institute (QBRI) was established as a member of Qatar Foundation by His Highness Sheikh Hamad Bin Khalifa Al Thani, Amir of Qatar, and Her Highness Sheikha Moza bint Nasser, and opened in November 2011. QBRI is seeking suitable candidates to fill the following vacancies:

- Principal Investigators
- Senior Scientists
- Scientists
- Postdoctoral Fellows
- Research Associates

The minimum requirements for these positions are a Master's Degree or PhD in Biomedical Sciences; hands-on laboratory experience; and a strong record of publications, conference papers, reports and/or technical contributions in any of the following research fields: **Cancer, Diabetes, Neurosciences, Infectious diseases, Cardiovascular diseases, Genetic diseases, Stem cells, Bioinformatics, Metabolomics, Proteomics, Genomics, and Molecular imaging.**

Compensation will include an attractive tax-free salary and additional benefits such as furnished accommodation, annual paid leave, medical insurance, etc.

Interested applicants should send their CVs to **recruitment@qf.org.qa** with the job title as the subject of the email. Only shortlisted applicants will be notified.

**About Qatar Foundation:**

Qatar Foundation for Education, Science and Community Development (QF) is a private, chartered, nonprofit organization, founded in 1995 by His Highness Sheikh Hamad Bin Khalifa Al Thani, Amir of Qatar. QF is guided by the principle that a nation's greatest resource is the potential of its people, and aims to develop that potential through a network of centers devoted to progressive education, research, and community welfare, thus supporting Qatar on its journey from a carbon-based to a knowledge-based economy.

**For more information please visit www.qf.org.qa**

RW227322R

---

# Communicate with Confidence

## How do you speak so others want to listen? Create an impact? Network like a pro?

VOX provides practical, imaginative training that boosts people's performance and prospects by strengthening their communication skills. We've coached researchers, academics and vice-chancellors at 25 universities, as well as staff in government departments and leading companies.

## Catch us...

**Catch Pete Bailie's session on Networking at the Naturejobs Expo, or email info@voxcoaching.com**

**VOX Coaching**

**www.voxcoaching.com**

U227333R

# :insideview
profile feature

**Jean-Claude Schmit,** *Director, Centre de Recherche Public de la Santé (CRP-Santé)*

*Luxembourg, one of Europe's smallest countries, has out-sized ambitions for science. Its plans to tailor its biomedical research into a nation-sized laboratory for personalized medicine is generating opportunities for scientists at all levels. Since the recent inception of a national program on biomedical research, scientists who join now will have the freedom to build both their laboratory and research programs from the ground up.*

*New CRP-Santé scientists will have lots of building materials. The country's excellent biomedical records and single-payer health insurance system provides robust data for both the science and economics of public health. The youth of the program means it can create biomedical research programs without being burdened by history. Luxembourg's sound economy and commitment to increasing biomedical research offers stability. And its combination of classic life sciences research (akin to the US National Institutes of Health) with public health research (akin to the US Centers for Disease Control and Prevention) allows for a more integrated approach to biomedical research — one that would be difficult to pioneer in either larger countries or countries with more established research programs. Jean-Claude Schmit, Director of the CRP-Santé, tells* Nature *why the new program is attractive to scientists seeking to draw their career on a blank page.*

### Q: What is the strongest attraction for scientists to join CRP-Santé?

The opportunity to build something from scratch is very attractive to some people. That is difficult to do when you go to a more established place. It is a higher risk, but on the other hand it's much more fun. Researchers opting for Luxembourg must accept the challenge of abandoning a "classic career" in one of the well established research settings abroad and take the "risk" (and the opportunity) to build something new.

### Q: What is the advantage of combining public health research with biological research?

We start from an "empty paper" because we do not have well established institutes with their own history in a specific field. Therefore we can be open-minded in envisioning a cross-competence collaboration. Despite the duel research functions, we can be fairly focused. We will have different types of research activities — basic, applied, public health-oriented — but in limited numbers of programs — all working close together in one physical place.

This will be a huge advantage in developing personalized medicine. To do so, you need input from research laboratories and the ability to test those approaches in patients—with excellent communication between the two fields. It also helps to dedicate your basic research laboratories toward experimental clinical needs. We have done so by building a strong clinical proteomics platform that we believe that will drive the future of personalized medicine, perhaps more so than genomics alone.

Finally, combining both functions allows you to look at real-world application of both basic and clinical findings. We will be able to model how different approaches to personalized medicine will affect the health care system as a whole, and determine which approaches are economically feasible.

> “we have the possibility to go much deeper in our analysis by using smaller samples of extremely well characterized data”

### Q: What are some of Luxembourg's strengths in public health research?

We are currently developing longitudinal cohorts we are following over time. That way we can look at chronic illnesses like Type 2 diabetes in combination with other life style diseases, like obesity, and cardiovascular diseases. Because we have very complete medical records on most citizens, we have the opportunity to have a rather detailed picture on what is happening in our country. So even though we cannot compete with large countries on the size and quantity of public health data sets, we have the possibility to go much deeper in our analysis by using smaller samples of extremely well characterized data. Other countries follow large populations over time as well, but in Luxembourg, it is done in one place, so there is no need to compete with other organizations for that information.

### Q: What is the role of health economics in your program?

The first question if you are speaking about actually applying innovation in health care is always, 'what will this cost?' For example: you design a new prognostic test in cancer, where you can now define a priori which person will respond to a given chemotherapy. Of course, this will increase costs (as novel diagnostic tests using genomics/proteomics are expensive). But using such tests will help you provide better care to the patient (by reducing side effects from useless medication or by finding a medication that will provide better response results) and use your chemotherapy in a way more targeted to how particular patients will respond to it. The challenge in applying health economics to personalized medicine involves making many such cost/benefit comparisons. In Luxembourg, such calculations are easier, because our health care system's costs are well known and our medical records are both centralized and standardized.

### Q: How has Luxembourg changed since it opened the new research priority in biomedical research?

There are clearer roles for the different actors in research now. CRP-Santé, the Integrated Biobank of Luxembourg, the University of Luxembourg, the National Health Laboratory and other actors are working much more closely. We now have a consortium of personalized medicine, which coordinates the interactions between these individual organizations. This allows us to exchange information in a simpler way. We have brought these actors together to discuss research strategy on a national level. Since Luxembourg is such a small country, we can have all the key players — less than 10 people — sitting around the table. Our system is big enough to be relevant, to serve as a model system, but small enough to act in a limber, agile manner. It is also a pilot scale for national personalized medicine. If it works in Luxembourg, then in a second step you could go to a bigger system, like Germany or France.

ADVERTISER RETAINS SOLE RESPONSIBILITY FOR CONTENT

# Excellence in Biomedical Research at the heart of Europe

## What we do:

We are a public sector organisation for basic, pre-clinical and clinical research in life sciences in Luxembourg. Our mission is to deliver scientific, economic and societal value for Luxembourg by performing research, studies and developments in the fields of clinically-oriented biomedical research and public health.

At the forefront of biomedical research, we aim at translating knowledge into clinical applications impacting on healthcare while shaping personalized medicine.

Our research activities are developed in the following thematic areas:

- Translational Cardiovascular Research
- Immunology
- Oncology
- Public Health
- Virology, Allergology and Immunity

These thematic areas receive support from 3 competence centers:

- The Luxembourg Biomedical Research Resources (LBR2): "omics" technologies
- The Competence Center for Methodology and Statistics (CCMS)
- The Clinical and Epidemiological Investigation Center (CIEC)

For the personalized medicine program, CRP-Santé collaborates with:

- TGen Foundation in Phoenix, Arizona
- Fred Hutchinson Cancer Research Center in Seattle, Washington
- Institute for Systems Biology, Seattle, Washington

## What we offer:

CRP-Santé offers the opportunity to evolve in an international environment, to join dynamic teams driven by internationally recognized leaders in biomedical research and to use state-of-the-art technology while developing continuously your personal and professional skills.

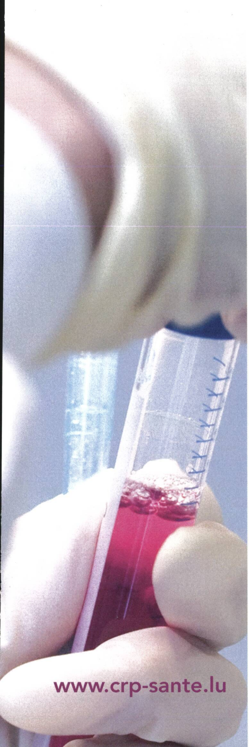
## Who we are currently looking for:

Research scientists in the field of virology or biostatistics, postdoc researchers with competences in bioinformatics or proteomics, postdoc researchers interested in the field of Cardiovascular disease or Toxicology, Clinical Research Associate.

## How to join us:

Learn more about CRP-Santé, job opportunities and fellowship program on our website: www.crp-sante.lu/jobs

For further details, please contact: recruitment@crp-sante.lu



hr
HR EXCELLENCE IN RESEARCH

CRP-Santé has been awarded the logo
"Human resources Excellence in Research"
by the European Commission

crp
SANTÉ
CENTRE DE RECHERCHE PUBLIC

www.crp-sante.lu

ALWAYS SEARCHING FOR MORE

naturejobs
CAREER
EXPO
LONDON 2012

# 20 SEPTEMBER 2012

## BUSINESS DESIGN CENTRE, LONDON

## CONFERENCE TOPICS AND SPEAKERS

**Keynote Lecture: Piled Higher & Deeper**
Jorge Cham, PHD Comics

**Transferable skills to progress your career (It's not enough to be a geek)**
Anna Price, King's College London
Stephen Tarling, University of Southampton
Sara Williams, Cardiff University

**Social media and managing your online presence**
Nicola Osborne, EDINA

**Networking**
Pete Bailie, VOX Coaching

**Funding through collaboration with industry**
Mark Christie, King's College London

**Open innovation – opening the kimono to a crowd**
John Lucas, Boehringer Ingelheim Pharma GmbH & Co. KG
Simon Schneider, InnoCentive

**Philanthropy and crowd-funding**
Simon Vincent, Cancer Research UK

**The essential guide to publishing in high-impact journals: 5 tips for success**
Nick Campbell, Macmillan Science Communication

**The european research area (ERA): what is in it for researchers?**
Anna Karaoglou, European Commission

**Best practice - how to be a grant go getter**
Stefan Bilbao, University of Edinburgh

**Singapore**
Paul Sharratt, ICES

**Studying in China: a French postdoc perspective**
Romain Amoit, University Claude Bernard Lyon 1

**Carry out research in Germany! – opportunities with funding**
Andreas Hoeschen, DAAD London
Svenja Rausch, DAAD London
Owen Vaughan, University of Liverpool & DAAD Alumnus

**Metrics: assessing and ranking excellence and impact**
Chair: Nick Campbell, Macmillan Science Communication

**What do researchers do?**
Alison Mitchell, Vitae

**Clinical drug development: understanding the career opportunities in the CRO industry**
Elizabeth Allen, Quintiles Drug Research Unit at Guy's Hospital

**Sustainability – a challenge for scientists**
Helen Sneddon, GlaxoSmithKline R&D Ltd

**Science policy and programme management (more exciting than it sounds)**
Simon Cutler, BBSRC

**Create your own job: the joy and pain of being an entrepreneur**
Ulf Grawunder , NBE-Therapeutics LLC

**From academia to industry and points in-between**
John Campbell, Miltenyi Biotec Ltd

**Register for the conference at www.nature.com/natureconferences/njce2012**

Gold Sponsors:
QBRI Qatar Biomedical Research Institute
crp SANTÉ CENTRE DE RECHERCHE PUBLIC
Research in Germany Land of Ideas

Silver Sponsors:
VIB
Otsuka
BASF The Chemical Company
WORK IN DENMARK
MERCK

Register at www.nature.com/natureconferences/njce2012

Contact Singapore
contactsingapore.sg

naturejobs

nature publishing group npg

## challenges love solutions

We think there's a solution for every problem. That's why BASF researchers from all disciplines always work on innovations with passion. Help us, in a modern environment, to find not just products but comprehensive solutions for tomorrow's challenges. That's how we create chemistry. At BASF. Find out more now and apply at: **www.basf.com/career**

**Get in touch**
BASF Services Europe GmbH
Recruiting Services Europe
P.O.Box 110248
10832 Berlin, Germany
Phone: 00800-33 0000 33
E-Mail: jobs@basf.com
**Career site: www.basf.com/career**
**Facebook: www.facebook.com/basfcareer**

BASF
The Chemical Company

W227292R

## Research is still your favorite?

## MAKE GREAT THINGS HAPPEN

**Opportunities for research scientists and professionals:** You have always strived to improve the patient's lives sustainably. Welcome to Merck. Contribute to exciting research projects in one of our high-growth therapeutic areas – cancer, neurodegenerative as well as autoimmune and inflammatory diseases, infertility and hormone disorders. We offer excellent development perspectives and exceptionally quick decision-making processes you will only find in mid-sized companies like ourselves. Join us and take part in shaping our diversified bio pharma business. Ready to tread new paths?

come2merck.com

MERCK

W227395R

naturejobs.com

## NEW - The *Naturejobs* Interactive Zone

A new addition to the Career Expo for 2012, take some time to search *Naturejobs* in our interactive zone. Search over **10,000+ vacancies** and read the latest **scientific recruitment news** and **career advice**

Located on the mezzanine level.

**www.naturejobs.com**

naturejobs CAREER EXPO LONDON 2012

**Follow us on:**

nature publishing group npg

Otsuka
Otsuka-people creating new products for better health worldwide

## Global Staff Scientist and Clinical Research Associate Positions Available

Otsuka Pharmaceutical Co., Ltd and its group companies (http://www.otsuka.co.jp/en/) invite applications for the positions of scientist & clinical research associate to lead the global research and development of new drugs. Successful applicants will lead a team of passionate professionals who strive to advance science, technology, and medicine by developing products designed to help people live better lives.

Successful applicants will start employment as a permanent employee of Otsuka Pharmaceutical Co., Ltd or its group companies and be stationed at Otsuka facilities worldwide including Japan, USA, EU, China, and so on. The starting date is as soon as practicable in 2012.

| | |
|---|---|
| **Disease areas:** | **Scientists -** CNS, Circulatory, Digestive, Cancer, Infection, Immunity, Ophthalmology |
| | **Clinical Research Associates -** Cancer |
| **Experiments:** | **Scientists -** PhD degree with two to five years combined post-doctoral and industrial experience, preferably in the pharmaceutical industry or biotechnology sector. |
| | **Clinical Research Associates -** Extensive experiences of anti-cancer drug development and expertise of regulatory registration in Europe. |

Applicants should include curriculum vitae written in English, a list of publications, and a statement explaining former research/work experience. Please keep application material to no more than five pages.

Applications should be sent to: **Email: recruitRD@otsuka.jp**
*(All the applications should be converted to PDF.)*

JP227318R

naturejobs
CAREER
EXPO
LONDON 2012

# 20 SEPTEMBER 2012

## BUSINESS DESIGN CENTRE, LONDON

## WORKSHOPS

### CRP-Santé Luxembourg: Excellence in Biomedical Research at the heart of Europe

Presented by CRP-Santé

CRP-Santé is a leading public research organization for basic, pre-clinical and clinical research in biomedical sciences located in Luxembourg, at the heart of Europe.

Our activities are focused on translational cardiovascular research, Immunology, Oncology, Virology, Allergology and Immunity, as well as Public Health.

We offer excellent quality-of-life benefits, including competitive salaries and generous benefits, such as access to a good healthcare system, training / personal development, and generous annual vacation leave.

### Academic & Research Careers with the Singapore University of Technology & Design

and

### Academic & Research Opportunities at Nanyang Technological University and the Lee Kong Chian School of Medicine

Presented by Contact Singapore

Ever thought about advancing your research career in Asia? Situated in the heart of Asia, Singapore is brimming with opportunities for bright and ambitious talent across varying levels of experience. We warmly welcome individuals who are looking for faculty positions or research openings in Singapore to attend.

### Identifying and Evidencing your Skills

Presented by Vitae

Writing job applications and CVs is really an exercise in convincing a prospective employer that your skills match the ones they are looking for. This workshop will explore how to really understand your own skills, and how they match up with what employers want. We will also cover how to evidence your abilities in order to convince the prospective employer of your suitability for the job.

### A Successful Research Career at Karolinska Institutet

Presented by Karolinska Institutet

Would you like to make a difference? Are you a creative person? Do you want to be inspired and work with like wised? Welcome to Karolinska Institutet!

In this workshop you will find out about research careers at one of the foremost medical universities in the world and what kind of support we offer during your research career. People from different support units will talk about doctoral education, supervision education, research funding, recruitment and career possibilities.

### Europe for Research Careers

Presented by European Commission

The workshop presents European Union's funding opportunities and services for first-class, up-and-coming researchers, who are keen to develop and fast-track their careers in Europe.

**Sign up for free and attend the workshops at no additional cost - visit www.nature.com/natureconferences/njce2012**

Gold Sponsors:

QBRI
crp SANTÉ CENTRE DE RECHERCHE PUBLIC
Research in Germany Land of Ideas

Silver Sponsors:

VIB
Otsuka
BASF The Chemical Company
WORK IN DENMARK
MERCK

Register at www.nature.com/natureconferences/njce2012

Contact Singapore
contactsingapore.sg

naturejobs

nature publishing group npg

# Boost your career in life sciences

G
C
C
G
T
G
1011
1011
G T C
A
T

**VIB, the Flanders Institute for Biotechnology,** is excellent in distinct research areas including plant biology, neurobiology, cancer, cardiovascular medicine, inflammation, bioinformatics, systems biology, proteomics and structural biology.

VIB is permanently on the look-out for enthusiastic scientists. Visit us at the Naturejobs Career Expo 2012 and enjoy Belgian chocolates, or connect with us at

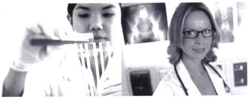
**www.vib.be/jobs**

VIB

W227334R

**omics@VIB postdoc call: 14 positions - deadline 30/09/2012**

---



**Would you like to explore career opportunities in a challenging and progressive Danish Life Science environment?**

**Please visit our stand and have a chat with recruitment consultants from Workindenmark, Technical University of Denmark and University of Copenhagen.**

UNIVERSITY OF COPENHAGEN

EURES

Technical University of Denmark DTU

**WORK IN DENMARK**

W227321R

---

naturejobs
CAREER EXPO | LONDON 2012

## 20 SEPTEMBER 2012

### BUSINESS DESIGN CENTRE, LONDON

The *Naturejobs* Career Expo in London is the UK's largest fair and conference focused exclusively on the scientific world.

**FREE Exhibition** featuring over 50 employers
**Insightful Conference** with high profile speakers
**Company Workshops** offering in-depth views

For more information visit:
**www.naturejobs.com/careerexpo**

bioscience careers
Advice on careers in and from biology

Event partner
Contact Singapore
contactsingapore.sg

naturejobs

nature publishing group npg

# naturejobs

**European Head Office, London** naturejobs@nature.com
The Macmillan Building, 4 Crinan Street, London N1 9XW, UK
Tel: +44 (0) 20 7843 4961 Fax: +44 (0) 20 7843 4996

**Global Head:**
Nils Moeller,+44 (0) 20 7843 4953 n.moeller@nature.com
**Southwest UK/RoW:**
Emma Hare, +44 (0) 20 7843 4944 e.hare@nature.com
**Northeast UK/Ireland:**
Matthew Ward, +44 (0) 20 7014 4059 m.ward@nature.com
**France/Switzerland/Belgium:**
Julie Aramburu, +44 (0) 20 7843 4994 j.aramburu@nature.com
**Scandinavia/Spain/Portugal/Italy:**
Megan Ball, +44 (0) 20 7843 4973 m.ball@nature.com
**North Germany/The Netherlands/Eastern Europe:**
Kristin Doelling, +44 (0) 20 7843 4970 k.doelling@nature.com
**South Germany/Austria:**
Maria Kubalova, +44 (0) 20 7014 4084 m.kubalova@nature.com
**Natureevents:**
Matthew Clare, +44 (0) 20 7014 4015 m.clare@nature.com
**Business Development Manager, Middle East:**
Jon Giuliani, +44 (0) 20 7843 4922 j.giuliani@nature.com

**USA Head Office, New York** naturejobs@us.nature.com
75 Varick Street, 9th Floor, New York, NY 10013-1917
Tel: +1 800 989 7718 Fax: +1 800 989 7103
**US Sales Manager:**
Ken Finnegan, +1 212 726 9248 k.finnegan@us.nature.com

**Japan Head Office, Tokyo**
Chiyoda Building, 2-37 Ichigayatamachi, Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751, Fax: +81 3 3267 8746
**Asia-Pacific Sales & Business Development Manager:**
Yuki Fujiwara, +81-3-3267-8765 y.fujiwara@natureasia.com
**Business Development Manager, Greater China/Singapore:**
Gloria To, +852 2811 7191 g.to@natureasia.com

**India**
3a, 4th Floor, DLF Corporate Park, Gurgaon, 122 002, India
**Advertisement Sales:**
Vikas Chawla, +91 1242881057 v.chawla@nature.com

---

TGAC The Genome Analysis Centre™
BBSRC EEDA Greater Norwich Development Partnership

ifr Institute of Food Research

## Yeast Genomics

The Institute of Food Research (IFR) and The Genome Analysis Centre (TGAC) are collaborating on comparative analysis of yeast strains, particularly those of interest to industrial biotechnology. An experienced computational biologist is sought to lead the analysis of the sequences generated. The preferred candidate will have a proven record in genome assembly and annotation. Expertise in comparative analysis of large-scale genomic datasets relating to eukaryotic organisms is essential. Candidates with experience in yeast genomics and/or an interest in genome analysis relevant to food and energy security are particularly encouraged to apply. Appointments will be based upon a track record of scientific excellence and strategic vision for the future. The post will be funded for 4 years in the first instance.

Salary on appointment will be dependent on qualifications and experience.

For further information and details of how to apply, please visit our web site http://jobs.ifr.ac.uk or contact Human Resources, Norwich BioScience Institutes Partnership, Norwich Bioscience Institutes, Norwich, NR4 7UH, UK, 01603 450149 quoting reference 1002250.

The closing date for applications will be 1st October 2012.

*The Institute of Food Research and The Genome Analysis Centre are Equal Opportunity Employers.*

U228213R

---

Memorial Sloan-Kettering Cancer Center
*The Best Cancer Care. Anywhere.*

# FACULTY POSITIONS

## Center for Stem Cell Biology & Developmental Biology Program

The Center for Stem Cell Biology (CSCB) and the Developmental Biology Program are seeking innovative individuals who have strong records of research accomplishments in stem cell biology for tenure-track positions at the Assistant Member level (equivalent to Assistant Professor). The CSCB was established to further the interaction of researchers with active stem cell research programs at the Center, to enhance and build novel stem cell related resources and to recruit leading-edge stem cell faculty. The CSCB is interested in research in any aspect of stem cell biology, including the study of stem cells in model organisms, tissue stem cells and the biology and application of human embryonic stem cells or induced pluripotent stem cells. Candidates with a strong interest in applying developmental approaches to studying stem cell biology are particularly encouraged to apply.

Applicants should have a PhD and/or MD degree, a productive postdoctoral experience, dedication to important problems in stem cell biology, and the potential to develop a strong independent program in stem cell research. Successful candidates will be appointed in the Developmental Biology Program and will become a member of the CSCB. Successful candidates are also eligible to hold additional appointments at the Gerstner Sloan-Kettering Graduate School of Biomedical Sciences, the Weill Cornell Graduate School of Medical Sciences, as well as the Tri-Institutional MD/PhD Training Program.

Successful applicants will have access to outstanding resources including expanded state-of-the-art facilities for human pluripotent stem cells, high-throughput chemical and genetic screening, mouse genetics, genomics, flow cytometry, advanced imaging and cancer biology. Memorial Sloan-Kettering offers a unique and exciting multidisciplinary research environment. Faculty in the CSCB participate in the Tri-Institutional stem cell initiative (with Weill Cornell and Rockefeller University) that links stem cell researchers across neighboring institutions.

The deadline for applications is **November 1, 2012**. Interested candidates should visit **facultysearch.ski.edu** to access the on-line faculty application. The site contains important information on the required application materials as well as deadlines for submission of letters of reference.

Inquiries may be sent to **Tiffany Lennon** at **lennont@mskcc.org**; to **Dr. Lorenz Studer, Director CSCB**, at **studerl@mskcc.org**; or to **Dr. Kathryn Anderson, Chair Developmental Biology Program**, at **k-anderson@ski.mskcc.org**.

**facultysearch.ski.edu**

MSKCC is an equal opportunity and affirmative action employer committed to diversity and inclusion in all aspects of recruiting and employment. All qualified individuals are encouraged to apply.

NW228173R

TGAC
The Genome Analysis Centre™

BBSRC EEDA East of England Development Agency Greater Norwich Development Partnership

# Building Excellence in Computational Bioscience

The Genome Analysis Centre (TGAC) is a vibrant research institute focused on the application of state of the art genomics and bioinformatics to advance plant, animal and microbial research and promote a sustainable bioeconomy. It is located with five other partners on the Norwich Research Park, a new bioscience cluster with an excellent reputation in plant and microbial sciences, interdisciplinary environmental science and food, diet and health. As a result of funding from BBSRC to support the growth of a national centre promoting bioinformatics for applied genomics, we are looking for exceptional candidates to fill senior posts within the Institute:

- **Head of Training and Outreach, Ref 1002282,** as a PhD scientist with experience in molecular biology or bioinformatics the post holder will lead the training and outreach programmes and teams for TGAC. The role will be key to communicating the Centre's capabilities, technologies and experience; training biologists / geneticists / breeders / virologists in the data analysis tools for today and the future; and supporting the national agenda to promote mathematics and computer science in biology.

- **Knowledge Exchange and Commercialisation Manager, Ref 1002286,** as a graduate from either a biological or life sciences background with an MBA or equivalent business qualification/experience the post holder will lead, develop further and implement TGAC's Knowledge Exchange and Commercialisation programme. The successful candidate will have the demonstrable ability to operate at a senior level, be able to represent TGAC in negotiations with commercial and public sector bodies and successfully manage and deliver large projects.

- **Head of Genomics, Ref: 1002281,** as a senior PhD scientist with demonstrated leadership of a genomics research programme the post holder will lead the scientific groups within our Genomics Department. The role will be key to reinforcing TGAC as a world-class national capability facility for genomics and bioinformatics in the UK. The successful candidate will have significant experience of genomics data analysis evidenced through a good publication record and successful grant applications.

- **Library Construction Group Leader, Ref 1002284,** as a PhD scientist with knowledge of molecular biology techniques the post holder will lead the Library Construction Group within the Genomics Department. The successful applicant will be a recognised expert in library construction, with demonstrable experience of managing a team and successful planning of multiple workflows. They will be highly motivated, an excellent communicator and able to manage change.

- **Sequencing Programme Manager, Ref 1002285,** as a science graduate with excellent up to date knowledge of DNA sequencing techniques and data processing the post holder will lead the Sequencing Programme Group within the Genomics Department. The successful applicant will have experience of managing a pipeline production in a scientific field and have worked in a customer/collaborator focused delivery environment.

- **Sequencing Informatics Group Leader, Ref 1002283,** as a PhD scientist with knowledge of bioinformatics and molecular biology the post holder will lead the Sequencing Informatics Group within the Bioinformatics Department. The successful applicant will have significant experience with computational processing or analysis associated with DNA sequencing technologies. They will also have a strong background in high-level programming languages, knowledge of data analysis and web development technologies.

For further information on TGAC, these opportunities and details of how to apply please visit: www.tgac.ac.uk or email Human Resources at nbi.recruitment@nbi.ac.uk. Closing date for all roles: 21st September 2012.

U228092R

SOUTHWESTERN
THE UNIVERSITY OF TEXAS
SOUTHWESTERN MEDICAL CENTER
AT DALLAS

# Assistant Professor in Human Genetics

The Center for Human Genetics (McDermott Center) at The University of Texas Southwestern Medical Center at Dallas invites applications for a tenure-track position of Assistant Professor. We are seeking individuals with innovative experimental research programs in human molecular genetics. Successful applicants will be expected to establish a vigorous independent research program and to teach students at the graduate level.

The individual should hold a graduate degree (MD, PhD or MD/PhD) and have completed a post-doctoral fellowship. The appointment will include a competitive salary, attractive start-up package, excellent laboratory space in a dynamic research environment with access to genetic core facilities. The faculty member will have a joint appointment in a basic science or clinical department. Applicants should submit their curriculum vitae containing a summary of past accomplishments, a statement of future objectives, and three professional references to:

Faculty Search Committee
c/o Susan Hayes
Dept. Administrator
McDermott Center for Human Growth & Development
UT Southwestern Medical Center at Dallas
5323 Harry Hines Boulevard
Dallas, Texas 75390-8591

Or by email to: Susan.Hayes@UTSouthwestern.edu

*UTSW is an Equal Opportunity Employer. Women and Minorities are encouraged to apply.*

NW227886R

---

Schulich
MEDICINE & DENTISTRY

# Tier 2 Canada Research Chair in Chemical Biology

Western

*Schulich School of Medicine & Dentistry and Faculty of Science Western University*

The Schulich School of Medicine & Dentistry and the Faculty of Science at Western University, one of Canada's leading research intensive universities, seek applicants for a Tier 2 Canada Research Chair (CRC) in Chemical Biology. In accordance with the regulations set for Tier 2 Canada Research Chairs (www.chairs-chaires.gc.ca/home-accueil-eng.aspx), the candidate will be an excellent emerging researcher who has demonstrated research creativity and innovation, and the potential to achieve international recognition in the field of Chemical Biology within the next five to ten years.

The Tier 2 CRC will be expected to establish an independent, externally funded research program in the area of Chemical Biology that will promote integration and synergy with existing areas of research strength in Proteomics & Protein Structure, Genomics & Bioinformatics, Synthesis and/or Materials & Biomaterials within the Schulich School of Medicine & Dentistry and the Faculty of Science at Western. Priority will be given to candidates with a strong record of productivity in chemical biology and interests in translational research. The candidate will have access to state-of-the-art facilities including the London Regional Proteomics Centre (www.lrpc.uwo.ca), the London Regional Genomics Centre (www.lrgc.ca) and the Western Nanofabrication Facility (http://www.uwo.ca/fab/). Furthermore, there will be excellent opportunities for collaboration with basic and clinical researchers at UWO and affiliated research institutes.

The successful applicant will hold a Ph.D. or an M.D., or equivalent, and will be a tenure track appointment at the position of Assistant Professor or at an Associate Professor level if qualifications and experience warrant. The appointment will be made to the Department of Biochemistry of Schulich School of Medicine & Dentistry and the Department of Chemistry of the Faculty of Science, with the opportunity for a cross-appointment to an appropriate Clinical Department, and an appointment as Scientist at the Robarts Research Institute and Lawson Health Research Institute.

With full time enrollment of about 32,000, The Western University graduates students from a range of academic and professional programs. Further information about the Schulich School of Medicine & Dentistry can be found at www.schulich.uwo.ca, the Faculty of Science at www.uwo.ca/sci and/or at www.uwo.ca. Western's Recruitment & Retention Office is available to assist in the transition of the successful applicant and his or her family.

The application should include a detailed curriculum vitae, a brief description of research accomplishments, representative publications, and the names of three references. As well, the candidate must include a proposal for an original, innovative, and high quality research program that would attract excellent trainees, students, and future researchers. Please send the complete application to:

Dr. Denise Figlewicz
Vice Dean, Research & Innovation, Schulich School of Medicine & Dentistry
Robarts Research Institute, Room 1240A
Western University
London, Ontario CANADA N6A 5C1
selection.committee@schulich.uwo.ca

Applications will be accepted until the position is filled. Review of applicants will begin after November 1, 2012.

*Positions are subject to budget approval. Applicants should have fluent written and oral communication skills in English. All qualified candidates are encouraged to apply; however, Canadians and permanent residents will be given priority. Western University is committed to employment equity and welcomes applications from all qualified women and men, including visible minorities, aboriginal people and persons with disabilities.*

NW227590R

# THE LAUFER CENTER
## FOR PHYSICAL AND QUANTITATIVE BIOLOGY
## AT STONY BROOK UNIVERSITY

**invites applications for**

## Two Endowed Assistant/Associate Professors in Physical, Quantitative or Systems Biology

**We seek two outstanding early- to mid-career researchers in physical, quantitative or systems biology for tenure-track positions at either the Assistant or Associate Professor levels. Research areas include, *but are not limited to*, computational, theoretical and/or experimental models of cellular networks, systems and evolution; systems pharmacology; or statistical biophysics. We seek the strongest candidates, irrespective of a particular research area within biology at the molecular or cellular levels.**

Each position focuses primarily on research, with minimal teaching, and is supported by a Laufer Endowed Chair that provides annual research support. The successful individual will have considerable freedom to choose his or her departmental affiliation.

Candidates should have a PhD and a strong record of research productivity. Please apply online at ***http://academicjobsonline.org***. Include a State employment application, cover letter, curriculum vitae, a two- to three-page description of research plans, and arrange to have a minimum of three letters of recommendation uploaded to the site. These positions will remain open until filled.

---

For more information or to obtain a State employment application, visit ***www.stonybrook.edu/jobs*** (Ref. #F-7314-12-06). Anticipated start date is September 1, 2013.

**THE LAUFER CENTER** is a center for research in physical and quantitative biology at Stony Brook University. In a newly renovated space on the Stony Brook campus, the Center provides a cross-disciplinary collaborative environment. The Center brings together Stony Brook researchers in chemistry, physics, applied mathematics and statistics, computer science, pharmacology, molecular genetics and microbiology, ecology and evolution; and researchers from Cold Spring Harbor Laboratory and Brookhaven National Laboratory. Stony Brook University is an exciting place for research, particularly in light of a recent major endowment, support from a New York State plan called NYSUNY 2020, and newly planned academic centers of excellence in bioinformatics, bioimaging and advanced computing.

LAUFER CENTER
FOR PHYSICAL AND QUANTITATIVE BIOLOGY

Stony Brook
University

STONY BROOK UNIVERSITY/SUNY IS AN AFFIRMATIVE ACTION, EQUAL OPPORTUNITY EDUCATOR AND EMPLOYER. 12080134

NW227913R

Accelerators | Photon Science | Particle Physics

Deutsches Elektronen-Synchrotron
A Research Centre of the Helmholtz Association

# DESY NANOLAB.

## DESY, Hamburg location, is seeking: Scientist, tenure track (m/f)

### DESY

DESY is one of the world's leading centres for the investigation of the structure of matter. DESY develops, runs and uses accelerators and detectors for photon science and particle physics.
The Nanolab provides DESY's synchrotron radiation users with complementary nanomaterial characterisation and synthesis methods. Further on, it promotes in-house research in the field of X-ray physics and nanoscience.

### The position

- Conception of large equipment for the Nanolab (e.g. focused ion beam, scanning electron microscope, e-beam lithography) and implementation in the context of the DESY photon sources
- Operation of Nanolab large equipment in collaboration with external users

### Requirements

- PhD in Physics, Material Science or Chemistry
- Well-founded knowledge of experimental techniques for nanocharacterisation and nanostructuring
- Well-founded, long-standing experience in scanning electron microscopy and/or focused ion beam techniques
- Profound knowledge of English
- Team player

For further information please contact Prof. Dr. Andreas Stierle, +49 40 8998-2005, andreas.stierle@desy.de.

The position is limited to 3 years.
Salary and benefits are commensurate with those of public service organisations in Germany. Classification is based upon qualifications and assigned duties. DESY operates flexible work schemes. Handicapped persons will be given preference to other equally qualified applicants. DESY is an equal opportunity, affirmative action employer and encourages applications from women. There is a bilingual kindergarten on the DESY site.

Please send your application quoting the reference code, also by email, to:
**Deutsches Elektronen-Synchrotron DESY**
Human Resources Department | Code: EM094/2012
Notkestraße 85, 22607 Hamburg | Germany
Phone: +49 40 8998-3392 | email: personal.abteilung@desy.de
**Deadline for applications: 30th September 2012**
**www.desy.de**

The Helmholtz Association is Germany's largest scientific organisation.
www.helmholtz.de

HELMHOLTZ | GEMEINSCHAFT

W228196R

---

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

MEDIZINISCHE FAKULTÄT

Die Medizinische Fakultät besetzt am Institut für Biochemie (Emil-Fischer-Zentrum) zum nächstmöglichen Zeitpunkt die

## W3-Professur (Lehrstuhl) für Biochemie und Molekulare Medizin

**(Nachfolge Prof. Dr. med. Cord-Michael Becker)**

Zu den Aufgaben gehört die angemessene Vertretung des Fachgebietes in Forschung und kurrikularer, insbesondere vorklinischer Lehre. Erwartet wird eine durch Veröffentlichungen und Drittmitteleinwerbungen nachgewiesene Exzellenz auf einem der etablierten Forschungsschwerpunkte der Fakultät, Mitarbeit in den Forschungsverbünden (Sonderforschungsbereiche, Graduiertenkollegs, Interdisziplinäres Zentrum für Klinische Forschung) und Zusammenarbeit mit den Gruppen am Institut, die Aspekte der neuralen Signaltransduktion und transkriptionellen Kontrolle bearbeiten. An der Medizinischen Fakultät bestehen die Studiengänge Humanmedizin, Zahnmedizin, Molekulare Medizin und Medical Process Management.

Einstellungsvoraussetzungen sind ein abgeschlossenes Hochschulstudium der Medizin oder der Naturwissenschaften, pädagogische Eignung, Promotion und zusätzliche wissenschaftliche Leistungen. Letztere werden durch Habilitation oder gleichwertige wissenschaftliche Leistungen, die auch außerhalb des Hochschulbereichs erbracht sein können, nachgewiesen oder können im Rahmen einer Juniorprofessur erbracht worden sein.

Bei einer Berufung in das Beamtenverhältnis dürfen Bewerberinnen/Bewerber zum Zeitpunkt der Ernennung das 52. Lebensjahr noch nicht vollendet haben. Ausnahmen in dringenden Fällen kann das Staatsministerium für Wissenschaft, Forschung und Kunst im Einvernehmen mit dem Staatsministerium der Finanzen zulassen (Art. 10 Abs. 3 Satz 2 BayHSchPG). Die Universität Erlangen-Nürnberg besitzt das Berufungsrecht.

Die Universität Erlangen-Nürnberg trägt das Zertifikat ‚audit familiengerechte hochschule' und bietet Unterstützung für Dual-Career-Paare an. Sie strebt eine Erhöhung des Anteils von Frauen in Forschung und Lehre an und bittet deshalb Wissenschaftlerinnen nachdrücklich um ihre Bewerbung.

Die Universität Erlangen-Nürnberg vertritt ein Konzept der intensiven Betreuung der Studierenden und erwartet deshalb eine hohe Präsenz der Lehrenden an der Universität.

Schwerbehinderte Bewerberinnen oder Bewerber werden bei ansonsten im Wesentlichen gleicher Eignung bevorzugt eingestellt.

Bewerbungen mit den üblichen Unterlagen (tabellarischer Lebenslauf, strukturierte Liste der Veröffentlichungen gegliedert nach Originalarbeiten, Buchbeiträgen, Übersichtsartikeln; Drittmittelaufkommen sowie Liste der Lehrveranstaltungen und wissenschaftliches Curriculum, Informationen hierzu finden Sie bei Berufungen unter http://www.dekanat.med.uni-erlangen.de) in Schriftform und auf CD-ROM/per E-Mail, ggf. Sonderdrucke sowie Zeugnisse und beglaubigte Kopien von Urkunden werden bis zum **19.10.2012** an den Dekan der Medizinischen Fakultät der FAU Erlangen-Nürnberg, Östliche Stadtmauerstr. 30a, 91054 Erlangen erbeten.

Grundzertifikat 2008 audit familiengerechte hochschule

www.fau.de

W228192R

---

**Follow *Naturejobs* on Twitter**

For the latest career news: **twitter.com/naturejobs**
For all job listings: **twitter.com/naturejobs_feed**

twitter

“The additional benefit of naturejobs.com is the cooperating portals [*Scientific American* and *Spektrum der Wissenschaft*] — these have generated a noticeable increase of interest in our offer.”

**Katarzyna Ginter,**
**Jagiellonian University,**
**Kraków, Poland**

naturejobs

## Faculty Positions in Continental Margins and Coastal Environments

Scripps Institution of Oceanography (SIO) at UC San Diego (http://sio.ucsd.edu) is committed to academic excellence and diversity within the faculty, staff, and student body. SIO is a world renowned center of marine research with approximately 200 principal investigators leading research programs on all aspects of earth, ocean and atmospheric sciences.

### Coastal Processes #10-439:

We seek outstanding candidates with complementary interests in the physical oceanography of the coastal environment. Topics of interest are focused on the shelf, inner shelf, and near shore, including tidal inlets, estuaries, river mouths, beaches, and headlands. Some examples of physical interest are: surf zone processes; surf zone-inner shelf exchange; sediment, chemical, and biological transport; waves (internal and surface); and coastal response to sea level changes or human activity.

### Coastal Ecology #10-440:

We seek outstanding candidates with expertise in the biology and ecology of the coastal environment. Areas of interest broadly include the interactions of marine organisms with one another and with their physical and chemical environment. Research topics may include the ecology of benthic and pelagic organisms, ecological processes and regulatory mechanisms, population recruitment and connectivity, response to habitat changes and disruptions, bio(geo)chemistry and genomic adaptation. Research programs that integrate observational, experimental and modeling approaches in the field and laboratory are encouraged.

### Geological, Geophysical and Geochemical Studies of Continental Margins #10-441:

We seek outstanding candidates from throughout the Earth sciences with complementary skills targeting continental margin development, evolution, and physical/chemical exchange across a variety of spatial and temporal scales. An emphasis on the marine portion of the system is envisioned, but candidates investigating onshore aspects of margin processes that cross the shoreline will also be considered. We encourage applicants whose research bears on key processes at margins, including geochemistry, geophysics, dynamics, seismology, tectonics, structural geology, geodesy, petrology, geochemical exchange (high- and low-temperature), geomorphology, glaciology, and ice-margin studies.

All successful candidates will be expected to teach classes, supervise research at both the graduate and undergraduate level and contribute to leadership on issues of equity and diversity. The positions require a PhD degree and a competitive record of publication, as well as evidence of the ability to conduct and fund an active research program consistent with the opportunity to have done so at this career level. The successful candidates will also have demonstrated the highest standards of scholarship and professional activity, or for junior scholars to have the potential thereof. The department is interested in candidates who have demonstrated commitment to excellence by providing leadership in teaching, research or service towards building an equitable and diverse scholarly environment.

Our preference is for hiring at the level of assistant professor for all positions, however appointments at all ranks may also be considered. Salary will depend on the experience of the successful applicant and will be based on the UCSD pay scales. Completed applications received by October 15, 2012 will be assured of consideration. Each position requires separate submission. All applications and related materials must be submitted electronically via Academic Personnel On-Line RECRUIT at: (https://apol-recruit.ucsd.edu). Emailed applications will not be accepted. Applicants should provide a current CV, cover letter including descriptions of their teaching experience, research interests, state if authorized to work in the U.S., and the names of three potential referees, along with their complete institution address, email address, phone and fax numbers, and a separate personal statement summarizing past or potential contributions to diversity (see http://facultyequity.ucsd.edu/Faculty-Applicant-C2D-Info.asp and http://sio.ucsd.edu/Diversity/ for further information). Questions about submission of applications may be addressed to Leslie Costi, (lcosti@ucsd.edu).

UCSD is an Affirmative Action/Equal Opportunity Employer with a strong institutional commitment to excellence through diversity.

NW227624R

---

MACMILLAN
**SCIENCE COMMUNICATION**

nature

## Publisher

Macmillan Science Communication (MSC) is the scientific communications and custom publishing division of Macmillan Publishers Ltd., publisher of the international science journal *Nature*, *Scientific American* and the family of *Nature* journals (through Nature Publishing Group (NPG)). MSC has exclusive access to the editorial and production expertise of NPG and thus is able to create scientific communication solutions of the highest possible standard.

MSC is currently seeking to employ a Publisher, to be based in London, with responsibility for Macmillan Science Communication activities and staff in this office, the Nature Index project (an extension of Nature Publishing Index http://www.natureasia.com/en/publishing-index/) and *Nature's* supplements / sponsored project program and partnerships.

The Publisher will be MSC's business lead in London, managing the London-based MSC team and *Nature's* supplements team as well as building necessary resources for Nature Index.

This role will take responsibility for maximising and sustaining commercial performance of the diverse science communication projects fulfilled by this team. Included in these responsibilities are publishing projects (digital, print and multimedia), science communication training products, editorial consultation / editing products and consultancies in science communication and publishing. Experienced in science publishing / communication, the ideal candidate will have a higher degree in a relevant field, strong commercial drive and aptitude for figures, proven experience in maximising revenue potential, business development and budget control. They would also have experience of running projects, controlling costs and hitting deadlines, as well as strong interpersonal and influential communication skills, with effective negotiation and presentation skills.

Additional desirable characteristics include: experience in developing and delivering custom publications, and a thorough understanding of the scientific research community and its needs. Experience of working in the domain of bibliographic analysis of the scientific literature would be an added plus for working on the Nature Index project.

This role is key to the expansion of MSC and is a demanding and varied position that is both intellectually stimulating and challenging. The successful candidate will be a flexible, dynamic and resourceful self-starter with excellent organizational skills.

The position is full-time in our London office. Salary will be commensurate with experience.

Please upload your application to vacancy **NPG/086/12** on http://jobs.macmillan.com/, including a short paragraph stating why you would be suited to the job and your CV. You must include a current address and daytime phone number and e-mail address, your current salary, and the date you would be available to start work.

**Closing Date: Monday 10th September 2012**

nature publishing group npg

IN227663R

# nature materials

## Associate Editor

*Nature Materials* is a prestigious international monthly journal (Impact Factor 32.841) covering all aspects of materials science and technology. We have an exciting opportunity available for a materials scientist to join our editorial team as an Associate Editor working on all aspects of the journal.

We are particularly interested in applicants with expertise in condensed matter physics but we would welcome applications from outstanding candidates in any area of materials science.

The ideal candidate should have a PhD and preferably postdoctoral experience with a strong research record. The successful candidate will play an important role in determining the representation of their field in the journal, and will work closely with the other editors on all aspects of the editorial process, including manuscript selection, commissioning and editing of Reviews and News & Views, and writing for the journal. A key aspect of the job is liaising with the scientific community through laboratory visits and international conferences.

This is a demanding and intellectually stimulating position. Broad scientific knowledge and training, excellent literary skills and a keen interest in the practice and communication of science are a prerequisite. The successful candidate must, therefore, be dynamic and outgoing and have excellent interpersonal skills. The salary and benefits will be competitive, reflecting the critical importance and responsibilities of this position.

The new editor will join our team in our London office. The *Nature Materials* team is part of a dynamic editorial and publishing environment that also includes *Nature, Nature Physics, Nature Nanotechnology* and *Nature Chemistry.*

Applicants should upload the following to vacancy **NPG/084/12** on http://jobs.macmillan.com/ - a CV (including their class of degree and a brief account of their research and other relevant experience), a News & View style piece (800 words or less) on a recent paper from related literature, and a brief cover letter explaining their interest in the post and their salary expectations.

**Closing Date: Monday 17th September 2012**

nature publishing group npg

IN227557R

---

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

GRADUATE SCHOOL OF QUANTITATIVE BIOSCIENCES MUNICH

QB

## Lecturer for biochemistry
## Lecturer for physics/applied math

The newly established **Graduate School of Quantitative Biosciences Munich (QBM)** seeks two lecturers to support its educational mission. The School is designed to prepare young life scientists for the emerging era of quantitative, systems-oriented bioscience and offers an interdisciplinary, international PhD program that brings together diverse fields, from biochemistry and medicine to bioinformatics, experimental and theoretical (bio-)physics, and applied mathematics. A key goal is to enable students to conduct research at the interface of traditional disciplinary boundaries and to communicate effectively across the divide between experiment and quantitative theory. The two lecturers with complementary expertise will be a vital part of this effort, both through formal classroom teaching and through informal mentoring and coaching.

QBM is funded through the German Excellence Initiative and is embedded in a strong and varied research environment in the biosciences in Munich. The School is based both at the central campus of the Ludwig-Maximilians-Universität (LMU) in downtown Munich and at the life sciences campus in Munich-Grosshadern/Martinsried.

**Your tasks**

- Teach cross-disciplinary primer courses aimed at conveying relevant basics to students trained in different fields: experimental life sciences for students with a theoretical background; biophysics and mathematical modeling for biochemists/biologists; biostatistics and bioinformatics for all students,
- teach tutorials and lab practicals for an intensive interdisciplinary core course that covers major problems in bioscience from multiple perspectives,
- work with QBM PIs to develop and refine the educational curriculum of the School,
- mentor and advise students,
- assist in the selection of students (reviewing applications, interviewing) and organize educational activities,
- assist the administrative coordinator in the operation of the School.

**Your profile**

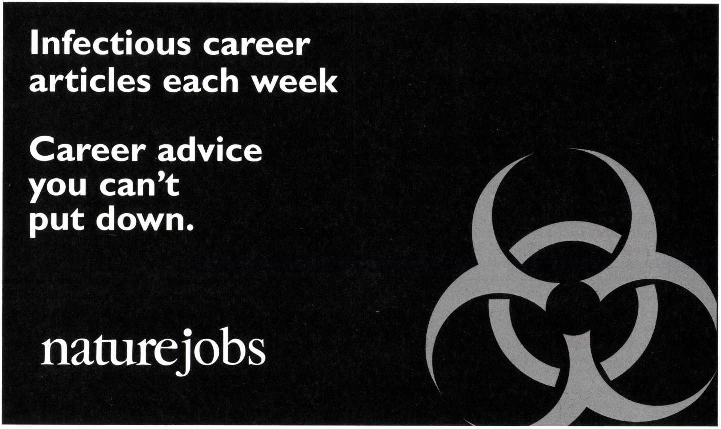
- PhD in biochemistry (experimental) or physics/applied mathematics (theoretical),
- postdoctoral and substantial research and/or teaching experience,
- you have the ability to elucidate key concepts in your field for non-specialists and enjoy working with students,
- you are interested in interdisciplinary research and communication,
- international experience and excellent English language skills,
- good organizational and interpersonal skills.

The salary is according to the TV-L (E 14). The positions are available immediately and will initially be for two years, but with the possibility of extension. The LMU is an equal opportunity employer; we particularly encourage the application of women. Disabled candidates with essentially equal qualifications will be given preference.

Please send a single pdf-file with your application (including cover letter, CV, degrees, references) by **September 30, 2012** to: **Prof. Dr. Ulrike Gaul (speaker of QBM), email: qbm-office@genzentrum.lmu.de,** http://www.qbm.lmu.de.

W227815R

---


Infectious career articles each week
Career advice you can't put down.
naturejobs

The FMI is an international biomedical research center with 300 members, including about 180 postdoctoral fellows and graduate students. It is affiliated with the University of Basel and is funded by the Novartis Research Foundation. The institute runs a successful international PhD program. Situated in Basel, Switzerland, the FMI offers an outstanding English-speaking scientific and cultural environment in the center of Europe.

**FMI**

Friedrich Miescher Institute
for Biomedical Research

# Group Leader Position, tenure track
# Cell fate determination and plasticity

The Friedrich Miescher Institute for Biomedical Research (FMI) invites applications for a tenure-track group leader position (equivalent to an assistant professorship) in its Mechanisms of Cancer program. We are seeking an outstanding individual who will establish a research program that explores **questions of cell fate using quantitative cell or tissue differentiation systems.** Competence in molecular mechanisms of developmental biology is welcome.

Opportunities exist for collaborative interactions with the other research teams in the institute with expertise in epigenetics, cancer and neurobiology. The FMI provides outstanding core facilities for experimental mouse genetics, live microscopy, cell sorting, genomics, protein crystallography, proteomics, and bioinformatics. Funding and salary are highly competitive.

Applications, including a CV, the names and emails of three referees, and a concise description of research interests and future plans should be submitted online at:

**www.fmi.ch/gl_search**

Informal inquiries can be sent to:
Dr. Susan Gasser (susan.gasser@fmi.ch)
4058 Basel, Switzerland

**The closing date for applications is November 30, 2012.**

W228134R

---

# *What's possible?*

gsk GlaxoSmithKline

## REALISE YOUR AMBITIONS ON OUR GLOBAL LEADERSHIP PROGRAMME

### Scientist - Esprit R&D Scientific Leadership Development Programme

Are you a **PhD-qualified scientist** eager to become an outstanding scientific leader? An inquisitive mind who would thrive on a diverse, intellectually demanding global leadership programme? Do you want to translate your scientific expertise into disease-beating medicines? And help people across the world to do more, feel better and live longer?

We are a science-led, global healthcare company that puts patients first. Our business is focused on developing medicines that improve the quality of human life. To meet the future strategic challenges this entails, we need to fast-track the finest scientific talent into senior leadership roles within our business.

Esprit R&D is a three-year, accelerated, multi-disciplinary and individually tailored programme. It is heavily backed by our senior management team, and it will give you incredible kudos as a biomedical scientist among your peers.

Over the three years, you'll undertake high-profile, global rotations that will give you in-depth biomedical knowledge and expertise. You'll apply this knowledge across a range of drug development disciplines from discovery research to commercialisation and business development. And as well as learning from thought leaders in your field, you'll benefit from unrivalled formal training, mentoring, coaching and development to enhance your leadership capabilities.

To be eligible for our Esprit R&D programme, you'll need:

- A PhD and completed at least one post-doc or a PhD and pharmaceutical or biotech experience gained in a highly credible organisation, plus excellent theoretical and practical skills.
- Commercial acumen and strategic decision-making skills.
- Innovation, creativity and the ability to solve complex problems.
- The ability to adapt and learn new areas while working as part of a team.
- The insight and curiosity to push boundaries in the biomedical field.
- Global mobility for a 12-month period.

Opportunities are available in the both the UK and US and will be open for applications from 3rd September to 24th September 2012. To find out more please visit www.gsk.com/esprit, or contact the Esprit R&D programme office on rd.esprit@gsk.com

Do more, feel better, live longer

U228209RL

**Massachusetts Institute of Technology**

**Come work with us!**

## Neuroscientist

The Picower Institute for Learning and Memory at the Massachusetts Institute of Technology is seeking outstanding neuroscientists for tenure track positions at the level of Assistant Professor.

We are interested in candidates studying development and/or plasticity of brain circuits at the cellular, circuit, and/or systems levels using a multi-faceted approach combining different methodologies and levels of analysis. We are particularly seeking, though not exclusively, candidates with strong cellular/molecular training.

Academic appointments will be in the Department of Brain and Cognitive Sciences. Candidates must have a commitment to excellence in undergraduate and graduate education, and must demonstrate the ability to develop a significant and independent research program.

**Applicants should submit a curriculum vitae, a summary of current and proposed research programs, an education plan, and arrange for three letters of recommendation. Applications are being accepted at Academic Jobs Online (https://academicjobsonline.org/ajo).**

**Search Contact: Prof. Elly Nedivi**

**Consideration of completed applications will begin immediately. Applications will be accepted for review until December 10, 2012.**

*MIT is an affirmative action employer, and we encourage applications from women and underrepresented minorities.*

**THE PICOWER INSTITUTE**
FOR LEARNING AND MEMORY

**http://web.mit.edu**

NW227613R

---

| le savoir vivant |

© Alexander Raths – Fotolia.com

### THE DOCTORAL SCHOOL OF THE FACULTY OF BIOLOGY AND MEDICINE, UNIVERSITY OF LAUSANNE, SWITZERLAND OFFERS

# phD fellowships in life sciences

starting october 2013

The Faculty of Biology and Medicine offers a broad choice of possibilities in different fields of biological, biomedical or public health research in the context of the outstanding scientific environment of Lausanne. More information about the Doctoral School, host laboratories, potential thesis directors and application procedure can be obtained at our website: **www.unil/fbm-phd.**

**Talented and motivated candidates who have or will have obtained Masters or equivalent degrees before Autumn 2013 are invited to submit their application before November 5th, 2012.**

The Phd program offers fellowships in laboratories of our Faculty.
The candidates may contact the thesis directors in their area of choice for information about specific projects.

For application procedure see: www.unil.ch/fbm-phd

W227437R

Unil
UNIL | Université de Lausanne
CHUV
Faculty of Biology and Medecine

---

### The University of Texas Health Science Center at San Antonio School of Medicine

### CHAIR, DEPARTMENT OF MICROBIOLOGY AND IMMUNOLOGY

The School of Medicine at the University of Texas Health Science Center at San Antonio seeks an individual with an outstanding record of scientific achievement, grant support, scholarly accomplishments and mentoring as its new **Chair** of the **Department of Microbiology and Immunology.** Dynamic leadership, communication and interpersonal skills, and keen vision are required. Reporting to the Dean of the School of Medicine and Vice President for Medical Affairs, the Chair will be responsible for continuing to build and maintain robust research programs, high quality graduate, medical and dental student education, and a collegial interactive culture. The UT Health Science Center is an Equal Employment Opportunity/Affirmative Action employer and is committed to excellence through diversity among its faculty, staff and students.

Candidates must be eligible for professorship with tenure. The ideal candidate will have a widely recognized national reputation in their field. The ability to foster a culture of collaboration, innovation, and accountability, across the Health Science Center is important. This is a wonderful opportunity for a visionary leader. UTHSCSA is a research intensive institution located in the Northwest region of San Antonio and sits as a gateway to the picturesque Texas Hill Country. San Antonio is a vibrant, dynamic, and multicultural city with much to offer, including an attractive cost-of-living.

Interested individuals should submit a letter of interest along with contact information for 3 references and a current CV electronically addressed to **Chris Walter, Ph.D., Chair of Cellular and Structural Biology and Chair of the Microbiology Search Committee c/o Shelly Evans, Assistant to the Dean, at elliss@uthscsa.edu.**

All faculty appointments are designated as security sensitive positions.

NW228156R

---

# FACULTY POSITIONS AT THE ROCKEFELLER UNIVERSITY

The Rockefeller University seeks exceptional, interactive, and creative scientists to join its faculty. We invite applications from outstanding candidates for tenure-track positions.

The University has a laboratory-based organizational structure that fosters interdisciplinary research. We encourage applications in the following areas:

- **Chemical & Structural Biology**
- **Genetics & Genomics**
- **Immunology, Virology & Microbiology**
- **Medical Sciences, Systems Physiology, & Human Genetics**
- **Molecular Cell Biology**
- **Neurosciences & Behavior**
- **Organismal Biology, Evolution, Ethology & Ecology**
- **Physical, Mathematical & Computational Biology**
- **Stem Cells, Development, Regeneration & Aging**

Details about specific subjects of research can be found at:
**http://www.rockefeller.edu/facultysearch.**

The Rockefeller University provides strong support for the research work of its faculty. The positions offer competitive salary, benefits and start-up funds, renovated laboratory space, access to state-of-the-art core facilities and extensive opportunities for collaboration both within the University and with neighboring institutions.

Applications are being accepted electronically through our **Online Application System** at **http://oas.rockefeller.edu**. Applicants should follow the online application procedure.

**The deadline for application submission is October 1, 2012.**

If you have questions regarding submitting an application, please contact our Administrator at **facultysearch@rockefeller.edu.**

The Rockefeller University is an Affirmative Action/Equal Opportunity/VEVRAA Employer and solicits applications from women and under-represented minorities.

NW227741RL

VE RI TAS

## Boston Children's Hospital
## Harvard Medical School

### Neuroscience & Developmental Neuroscience
### Assistant/Associate Professor

Applications are being considered for full-time, tenure-track research faculty positions at Boston Children's Hospital. The successful candidates will hold either a PhD or MD degree and will have research interests relevant to the development, function, or pathology of the nervous system, broadly defined, with an emphasis on cellular and molecular biology. The available faculty position is within the vibrant and highly interactive research environment of the F.M. Kirby Neurobiology Center directed by Clifford Woolf. This basic science research program, affiliated with the Department of Neurology, Children's Hospital, resides within a very strong and collegial research community in neuroscience and related disciplines throughout Harvard Medical School and its affiliated hospitals, and Harvard University (see http://childrenshospital.org/research/Site3038/mainpageS3038P0.html). Modern laboratory space is available in the new Center for Life Sciences Building. We are seeking exceptional scientists to establish vigorous independent research programs, form productive collaborations and, if appropriate, promote translation of their research. The investigators will hold both Boston Children's Hospital and Harvard Medical School faculty appointments. Please submit a current CV, a 2 or 3 page description of research interests and future directions, and three to five reference letters. Materials should be sent by November 30, 2012 to:

Neuroscience Search Committee c/o Maliha Chaudhry
bch.neurobiology@gmail.com
Equal Opportunity/Affirmative Action Employer

NW228170R

---

**Duke University**
**School of Medicine**

## Tenure-track Faculty Position
## (Assistant/Associate/Full Professor)
## Department of Molecular Genetics and Microbiology, Duke University Medical Center

Applications are invited for a tenure-track position in the Department of Molecular Genetics and Microbiology at Duke University Medical Center ( http://mgm.duke.edu/ ). We are seeking individuals that employ genetics to investigate important biological problems. There are no restrictions concerning the research area or the model organism used. Nonetheless, individuals working in the following areas are particularly encouraged to apply: 1) the genetic basis for behavior and disease in Drosophila, Caenorhabditis or mammalian models, 2) functional studies of non-coding RNAs, and 3) genetic analysis of genome stability in multi-cellular organisms.

The individual hired is expected to develop a strong, independent research program that will complement existing areas of expertise within the Department. Research strengths within the Department include RNA biology, DNA repair and recombination, microbial pathogenesis (viral, bacterial and fungal), and human/mammalian genetics.

Applications should include a curriculum vitae, a description of research accomplishments and plans for future research (please limit to three pages), and reprints of three representative publications. Interested applicants should submit application materials (cover letter, CV, statement of research accomplishments/interests and at least three references) for this job posting on AcademicJobsOnline at https://academicjobsonline.org/ajo/jobs/1676. Applications should be received by November 15, 2012.

Further inquiries may be directed to Rachel Mullis at rachel.mullis@duke.edu.

***The deadline for receipt of applications is November 15, 2012.***

*Women and minorities are encouraged to apply.*
*Duke University is An Equal Opportunity/Affirmative Action Employer.*

NW227755RL

---

MACMILLAN
**SCIENCE COMMUNICATION**

## Assistant Editor – Macmillan Science Communication

Macmillan Science Communication (MSC) seeks an assistant editor for a range of high profile scientific communication projects. You will be working within an expert team of scientific communications professionals on custom web and print publications produced on behalf of top scientific institutions in the Asia-Pacific region.

The job involves

- Ensuring that the English language content of articles is of high quality, readability and accuracy.
- Copy editing to house style of written content for MSC projects. This includes creating titles, standfirsts, image captions and other short pieces of text as required.
- Uploading content for web production using a content management system (CMS) and web proofing.
- Placing of content for print production using Adobe Indesign and print proofing.
- Ensuring that content is delivered and published to schedule.

The successful candidate will have a sound scientific background with at least a bachelor degree in science, or in a journalism or English-related discipline but with demonstrated fluency in science. The candidate will speak of English to a native level of fluency and will have demonstrable aptitude in scientific and/or technical copy editing. The candidate will receive full training in house styles and procedures. Outstanding written and oral communication skills and the ability to work to tight and changeable deadlines are required. Experience of using content management systems and Adobe Indesign would be an advantage.

Macmillan Science Communication (MSC) is the scientific custom publishing division of Macmillan Publishers Ltd. Macmillan Publishers Ltd. is the owner of Nature Publishing Group (NPG), publisher of the international science journal Nature, Scientific American, and a family of Nature-branded research journals. MSC has exclusive access to the editorial and production expertise of NPG (including Scientific American) in print, online and audio-visual media and thus is able to design and create content with the highest editorial and production standards.

The position will be based at the Asia-Pacific regional headquarters of Macmillan Science Communication in Tokyo, Japan.

Applicants should send a CV, and a brief cover letter explaining their interest in, and suitability for, the post and examples of any previous written work which may be of relevance to:

*Matthew Salter, Macmillan Science Communication (Asia-Pacific), Chiyoda Building, 2-37 Ichigayatamachi, Shinjuku-ku, Tokyo 162-0843 Japan or by email to: jobs@mscasiapacific.com*

**The closing date for applications is September 7th 2012**

IN227429R

PRINCETON UNIVERSITY

## ASSISTANT PROFESSORSHIP MOLECULAR EVOLUTION

Princeton University's Departments of Ecology and Evolutionary Biology and of Molecular Biology jointly plan to hire one individual at the level of tenure-track Assistant Professor, with a focus on molecular evolution. We seek an individual who would have strong connections in both departments. Sample areas might include, but are not limited to, microbial evolution, experimental evolution, epigenetics, metagenomics, and/or the evolution of development, using traditional and/or emerging model systems. We seek applicants who pursue experimental research that aims for significant conceptual and empirical integration across traditional disciplinary boundaries and who have a strong commitment to teaching.

Applicants should write a vision statement, no longer than 2 pages, that outlines one or more major unsolved problems in their field and how they plan to address them. The vision statement must be more than a précis of the applicant's prior and current research. Applications, including the vision statement, curriculum vitae, three reprints and contact information for three references, can be submitted online via http://jobs.princeton.edu, requisition #1200524. Screening of applications will begin 1 October, 2012.

*Princeton University is an equal opportunity employer and complies with applicable EEO and affirmative action regulations.*

NW228062R

---

BIOMEDICAL ENGINEERING
Carnegie Mellon

**The Department of Biomedical Engineering at Carnegie Mellon University, undergoing rapid expansion, seeks the appointment of tenure-track faculty members at all levels.**

A successful candidate is expected to build a vigorous research program that meshes with and enhances the extensive research network as described on http://www.bme.cmu.edu, particularly in the areas of cellular and cardiovascular biomechanics, biomedical imaging and signal processing, bioMEMS, and smart medical devices and materials. The College of Engineering at Carnegie Mellon is ranked consistently among the top 10 engineering schools by USNWR. Faculty members of the Department enjoy a strong university-wide collaborative culture, excellent research support, modest teaching load, outstanding students, and comprehensive benefits.

Applicants should send CV, three letters of recommendation, and vision statements on research and teaching to bme-search@andrew.cmu.edu. Applications will be reviewed on a rolling basis until February 2013.

*Carnegie Mellon University, an Equal Opportunity/Affirmative Action Employer, strongly encourages applications from women, under-represented minorities, individuals with disabilities, and individuals from disadvantaged backgrounds.*

NW228043R

---

UNIVERSITY OF MARYLAND

## Open Rank Faculty Position
## University of Maryland College Park
## Department of Cell Biology and Molecular Genetics

The Department of Cell Biology and Molecular Genetics at the University of Maryland College Park is seeking to fill a tenure-track faculty position in the broad areas of cell biology and molecular genetics using animal, plant, or microbial systems. The appointment may be made at the Assistant, Associate, or Full Professor level, commensurate with qualifications and experience.

The successful candidate will be expected to maintain a cutting-edge externally funded research program that synergizes with existing core groups in the department including genomics and gene regulation, RNA structure and function, plant biology, microbial pathogenesis and immunology, and virology. The appointed candidate will also participate in undergraduate and graduate teaching. Applicants must have a doctorate degree, an outstanding publication record, and a commitment to excellence in teaching.

The University of Maryland, College Park is the flagship campus of the University System of Maryland. Close proximity to Washington DC, Baltimore, and the Maryland Biotechnology Corridor facilitates interactions with an extraordinary range of major research institutions such as NIH, NIST, FDA, USDA and JCVI, in addition to providing a rich cultural environment.

Applications should be submitted electronically to https://jobs.umd.edu/applicants/Central?quickFind=57676 position number 118570, and addressed to Dr. Zhongchi Liu, chair of the faculty search committee. Applications should consist of a single PDF file containing (1) a cover letter, (2) curriculum vitae, (3) summary of research plans (maximum two pages) and teaching philosophy (one page), and (4) contact information for at least three references. Complete applications should be received by October 15, 2012, but will be accepted until the position is filled.

*The University of Maryland is an Affirmative Action/Equal Opportunity Employer. Women and members of underrepresented groups are especially encouraged to apply.*

NW228225R

---

## FACULTY POSITION
## Princeton University

**The Molecular Biology Department at Princeton University invites applications for a faculty position at the Assistant Professor level. We are seeking an outstanding investigator to address fundamental questions in Host-Microbe Interactions.**

The University has a strong commitment to interdisciplinary studies, especially in the areas of systems biology, imaging, genomics, biophysics and evolution. The department has high-level computing and microscope facilities, DNA array and high throughput sequencing technologies, mass spectrometry, and state of the art vivarium. Applicants must have an excellent record of research productivity and demonstrate the ability to develop a rigorous research program.

All applicants must have a Ph.D. or equivalent degree and a commitment to teaching at the undergraduate and graduate levels. Applications must be submitted online at http://jobs.princeton.edu, requisition **#1200547** and should include a cover letter, curriculum vitae, a two-page research description, and contact information for three references. All materials must be submitted as PDF files.

Screening of applications will begin 1 October, 2012.

*Princeton University is an equal opportunity employer and complies with applicable EEO and affirmative action regulations.*

NW228073R

---

## Three Principal Investigator Positions, Bigelow Laboratory for Ocean Sciences

Bigelow Laboratory invites applications for three Principal Investigator positions associated with the three centers of research in the Laboratory's new 60,000 ft$^2$ Ocean Science and Education Campus. The targeted research areas for these positions are: 1) **Blue Biotechnology** (e.g. marine natural products chemistry/molecular biology, evolution of ocean microbial communities, algal biomass, biofuels, marine phycology/mycology); 2) **Ocean Biogeochemistry and Climate Change** (e.g. modeling, ocean observing, carbon biogeochemistry, controls on the ocean biological pump, paleobiology, microbiology of the cryosphere, stable isotope biogeochemistry, environmental radiochemistry), and 3) Ocean Health (e.g. systems microbiology, chemical ecology, benthic ecology, coastal impacts, plankton physiology, environmental chemistry). Cross-cutting themes for these positions include human impacts, foundations of marine food webs, and climate change.

Minimum requirements are a Ph.D. with two years of postdoctoral experience and the capability to acquire external funding and lead scientific programs. Bigelow Laboratory is a primarily soft money institution with the expectation that the PI will generate external funding and be involved in the Laboratory's institutionally-funded teaching and technology transfer activities. Salary and start-up packages are negotiable. The Laboratory provides generous medical/retirement benefits.

For more information and application details, see our Web Site: http://www.bigelow.org/news/position_openings/3pi_positions/ referring to Job # SRS-2012-13.

For full consideration, the application should be received by **26 October 2012**. The search will continue until the positions are filled but the goal is to have these positions filled by summer 2013.

*Bigelow Laboratory is an Equal Opportunity Employer.*

NW227976R

---

## FACULTY APPOINTMENT IN CELL BIOLOGY
## Department of Cell Biology
## Duke University Medical Center

The Department of Cell Biology at Duke University Medical Center is hiring a tenure-track faculty member. The preferred level is for an Assistant Professor who would start his/her independent career at Duke. Candidates with demonstrated expertise in stem cell biology, tissue regeneration, or animal cell structure/function are especially encouraged to apply. Preference will be given to candidates whose activities complement and strengthen research programs within the department. Successful candidates will have track records of creativity and productivity, demonstrated potential for outstanding future achievements, and a desire to participate in an exciting, interactive, and growing community. They will be expected to mount a productive and innovative research program, to obtain outside funding, and to participate actively in graduate training. Candidates must have a Ph.D., M.D. or equivalent degree. Qualified minority candidates are especially encouraged to apply. For full consideration, application should be submitted by December 1, 2012, although applications will be considered until the position is filled. Applicants should submit a curriculum vitae, a brief summary of current and proposed research, and 3 letters of recommendation to:

**https://academicjobsonline.org/ajo/jobs/1810**

Questions may be directed to Dr. Brigid Hogan, Chair, Department of Cell Biology, (brigid.hogan@duke.edu)

*Duke University is an Equal Opportunity Affirmative Action Employer.*

NW228174R

CCRCB
Centre for
Cancer Research
& Cell Biology

## Director

Ref: 1104

### Centre for Cancer Research and Cell Biology

One of the leading universities in the UK and Ireland, Queen's University Belfast is continuing to strengthen its world-class reputation and the international impact of its biomedical research by seeking to recruit a new Director of the internationally renowned Centre for Cancer Research and Cell Biology (CCRCB).

We are seeking an exceptional individual who has the vision, drive and outstanding leadership skills to further develop the CCRCB's world-class research programme that spans basic through to translational to clinical research. The appointee will also ensure that the IPR arising from the CCRCB research programme is exploited commercially and will further develop the entrepreneurial ethos of the Centre which has already produced successful spin-out companies employing staff in both Northern Ireland and in the USA.

The CCRCB is one of four leading Research Centres which, along with the Wellcome Clinical Research Facility, make up the University's Institute of Health Sciences. The CCRCB's mission is to drive discovery, innovation and the development of targeted therapies in order to accelerate the defeat of cancer. It is part of the Experimental Cancer Medicine network and is a CR-UK designated Centre. The CCRCB is a cross-faculty, interdisciplinary research centre, with over 250 clinical and basic researchers from across the world. It has established international partnerships with leading global institutions such as the US National Cancer Institute. Most recently, in recognition of the Centre's leadership of the Northern Ireland Comprehensive Cancer Services (CCS) programme, along with the Northern Ireland Clinical Cancer Centre and the Northern Ireland Cancer Registry, the University was presented with the Diamond Jubilee Queen's Anniversary Prize.

The closing date is 12noon Monday 22 October 2012.

For further information and for details on how to apply, please visit our Academic Recruitment website at **www.qub.ac.uk/jobs** or contact Perrett Laver, our appointed executive search partner, at 020 7340 6200 or at qub@perrettlaver.com.

The University is committed to equality of opportunity and to selection on merit. It therefore welcomes pplications from all sections of society and particularly welcomes applications from people with a disability.

THE QUEEN'S
ANNIVERSARY PRIZES
FOR HIGHER AND FURTHER EDUCATION
2011

**Personnel Department**
**Queen's University Belfast**
Belfast, BT7 1NN.
Tel (028) 90973044
Fax (028) 90971040
E-mail on personnel@qub.ac.uk

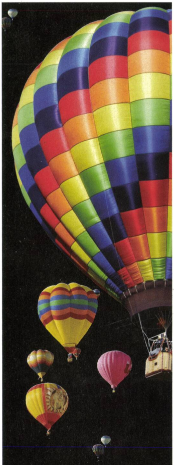
Perrett
Laver

Queen's University
Belfast

RUSSELL
INTERNATIONAL
EXCELLENCE
GROUP

**Queen's University Belfast is a member of the Russell Group of universities.**
**One of the United Kingdom's top 20 research-intensive universities.**

U228202R



HIGHER
& HIRE!

Post all your vacancies for FREE on the largest dedicated job board for the scientific community:

naturejobs.com

nature publishing group npg

"*Naturejobs* has been a great way to find the right people to build our research team. In less than 1 month, I received over 40 inquiries for my advertised postdoctoral position, with many promising candidates. As a result, I have hired two very well-qualified fellows and am excited about moving our research forward with our new team."

**Dr Kathleen Martin,**
**Yale School of Medicine, USA**

naturejobs

## THE CHINESE UNIVERSITY OF HONG KONG

Applications are invited for:-

## Department of Physics
## Assistant Professors
*(Ref. 1213/004(540)/2)*

The Department invites applications for faculty positions at the level of Assistant Professor with prospect for substantiation tenable from the academic year 2013-2014. The Department anticipates that there will be multiple positions.

Applicants should have a relevant PhD degree with postdoctoral research experience. Outstanding candidates in all areas of physics, particularly in the fields of experimental condensed matter physics, atomic, molecular and optical physics, and astrophysics, are welcome to apply. Successful candidates are expected to demonstrate a strong record of research accomplishments, potential for establishing a significant externally funded research programme, and a strong interest in teaching at undergraduate and postgraduate levels. Appointments will normally be made on contract basis for up to three years initially, which, subject to mutual agreement, may lead to longer-term appointment or substantiation later (substantive appointment may be considered during the second three-year contract). Applications will be accepted until the positions are filled.

**Salary and Fringe Benefits**
Salary will be highly competitive, commensurate with qualifications and experience. The University offers a comprehensive fringe benefit package, including medical care, a contract-end gratuity for appointments of two years or longer, and housing benefits for eligible appointees. Further information about the University and the general terms of service for appointments is available at *http://www.per.cuhk.edu.hk*. The terms mentioned herein are for reference only and are subject to revision by the University.

**Application Procedure**
Applications (comprising a full curriculum vitae, a detailed publication list with three selected published papers, a research plan, a teaching statement, and three letters of recommendation) should be sent to Professor Ke-Qing Xia, Chairman, Department of Physics, The Chinese University of Hong Kong, Shatin, Hong Kong (email: *physics@cuhk.edu.hk*; fax: (852) 2603 5204). The Personal Information Collection Statement will be provided upon request. Please quote the reference number and mark 'Application - Confidential' on cover.

JP228080R

---

University of Veterinary Medicine Vienna
Vice Rectorate of Research

vetmeduni vienna

## Postdoctoral Fellowships

Vetmeduni Vienna invites applications for to 2 post doctoral fellowships as early career opportunities for developing outstanding research programs.

Starting 2012 the university will award each fellowship a 4 year funding.

Salary is €47.340,- (before tax) and research expense is €10.000,- per year.

No teaching or administrative duties are associated with these appointments, however candidates are expected to apply for additional research funding.

Applications in the following areas are invited.
(For details please use link: http://www.vetmeduni.ac.at/postdoc/)

| Code: | Research Area | Principal Investigators |
|---|---|---|
| PD0112 | RNA interference (RNAi) as a possible control of whirling disease in rainbow trout (Oncorhynchus mykiss) | Mansour El-Matbouli |
| PD0212 | Early embryo development in the context of pathophysiology of reproduction and embryo-maternal crosstalk | Urban Besenfelder/Gottfried Brem |

**Closing date for applications: Sept 27th 2012**

A PhD degree in Life Sciences or Dr.med.vet. or Dr.med.degree is required.

Application should contain: full CV, publication list, application letter, 2 letters of recommendation and should be submitted electronically to: rektorat@vetmeduni.ac.at (Also for informal enquiries.)

Vetmeduni Vienna is committed to the recruitment and advancement of female scientists. Appointments will be subjected to a pre employment screening.

W228085R

---

UMASS LOWELL
*Learning with Purpose*

Careers with **Mass Appeal**

## Department Chair for Biological Sciences

The University of Massachusetts Lowell Department of Biological Sciences invites applications for a full-time tenured faculty position at the rank of Full Professor to serve as Department Chair starting January 2013 or as soon as possible thereafter. We are seeking a distinguished mid-career individual with the vision and leadership to promote a rapidly growing department and to provide leadership for the further development of research and academic programs.

The Department has over 500 students, offers BS and MS degrees, and students can pursue PhD degrees through several interdisciplinary programs. Current faculty research interests include genetics, neurobiology, cancer biology, invertebrate biology, developmental biology, structural biology, virology, microbial ecology, evolution, and biogeochemistry. Our campus is located very near the vibrant academic and commercial biotechnology centers of Boston, Cambridge and Worcester.

**Minimum Qualifications:** Two years' experience in leadership; five years' teaching at the undergraduate and graduate levels.

**Preferred Qualifications:** Significant record of funding and publication from a sustainable externally funded research program; excellent communication and interpersonal skills; ability to work with diverse student and faculty population.

**To apply, visit http://jobs.uml.edu. Please submit a curriculum vitae, copies of several recent research publications, a statement of research and teaching interests, not to exceed three pages, and three letters of recommendation. Review of applications will begin September 30, 2012, and continue until the position is filled; however, the posting may close when an adequate number of qualified applications is received. Thank you for considering the University of Massachusetts Lowell as an employer of choice. We look forward to receiving your application.**

*The University of Massachusetts Lowell is committed to increasing diversity in its faculty, staff, and student populations, as well as curriculum and support programs, while promoting an inclusive environment. We seek candidates who can contribute to that goal and encourage you to apply and to identify your strengths in this area.*

NW228055R

---

CHEMISTRY & CHEMICAL BIOLOGY

McMaster University

## Tenure-Track Assistant Professor Position in Chemistry of Biointerfaces

The Department of Chemistry & Chemical Biology at McMaster University solicits applications to fill a tenure-track position at the rank of Assistant Professor in the area of biointerfaces chemistry, effective July 1, 2013. Applications are encouraged in all areas of biointerfaces chemistry, including preparation, characterization, modeling or applications of biointerfaces. Scientists with research interests that have applications in the biosciences, environmental science and materials/nanoscience areas are encouraged to apply.

Our aim is to enhance interdisciplinary science at McMaster, a university with a longstanding history of collaborative research initiatives. In addition to having access to the equipment and facilities in this department, the successful candidate will also have access to the facilities in the Brockhouse Institute for Materials Research, the Canadian Centre for Electron Microscopy, the Centre for Microbial Chemical Biology and the newly established Biointerfaces Institute. This facility consists of several instrument suites for high throughput development and characterization of biomaterials and biointerfaces, including 850 MHz solid-state biological NMR, imaging MALDI mass spectrometry, imaging XPS and automated materials synthesis and screening infrastructure, with current research themes in bioactive surfaces for sensing and separation, ophthalmic biomaterials, and fundamental studies of biointerface interactions.

For additional information on the department, see: http://www.chemistry.mcmaster.ca Applicants should clearly demonstrate potential to develop a prominent, externally funded research program and to be committed to excellence in teaching at the graduate and undergraduate levels. Candidates must have a doctoral degree in chemistry, or a closely related field, postdoctoral experience and a promising record of research scholarship and productivity. Application materials must include a cover letter, curriculum vitae, a statement of teaching interests, and detailed descriptions of at least three research projects that exemplify your proposed research program. Please include a listing of the major instrumentation and equipment necessary to pursue each project. Review of applications will begin after November 1, 2012 and will continue until the position is filled. Please send these materials and arrange for three letters of recommendation to be sent to:

Dr. John D. Brennan
Department of Chemistry & Chemical Biology,
McMaster University,
1280 Main Street West,
Hamilton, Ontario L8S 4M1
CANADA

All qualified candidates are encouraged to apply; however, Canadian citizens and permanent residents will be given priority. McMaster University is strongly committed to employment equity within its community, and to recruiting a diverse faculty and staff. The University encourages applications from all qualified candidates, including women, members of visible minorities, Aboriginal persons, members of sexual minorities, and persons with disabilities.

NW228071R

JDRF IMPROVING LIVES. CURING TYPE 1 DIABETES.

# Improving Lives and Curing Type 1 Diabetes

**JDRF** has funded over $1.6 billion in type 1 diabetes (T1D) R&D to cure, better treat, and prevent T1D. This year we plan to launch new initiatives in multiple areas, such as beta cell regeneration, immune therapies, vaccines, biomarkers, sensors, and clinical translation. JDRF is seeking novel approaches to the challenges of T1D because our supporters depend on us—and we depend on you.

Read JDRF's mission and strategy in:
**Diabetes. January 2012;61:30-35**

Review JDRF's FY2013 research priorities at:
**jdrf.org/priorities**

Monitor current and future JDRF funding opportunities at:
**jdrf.org/funding**

NW228070R

---

FAS Center for Systems

VE RI TAS

## Harvard University
## FAS Center for Systems Biology
## Faculty Position

The Faculty of Arts and Sciences (FAS) Center for Systems Biology (http://sysbio.harvard.edu/csb) seeks outstanding candidates for a faculty position at the rank of assistant professor (tenure track). The Center, whose common research space houses faculty from a spectrum of academic departments and the Bauer Fellows, fosters interactions across disciplinary boundaries. We are interested in candidates involved in research in systems biology, using either experimental or computational approaches. A PhD in a relevant discipline is required.

The successful candidate will hold an academic appointment in a FAS natural science department such as, but not restricted to, Molecular and Cellular Biology, Organismic and Evolutionary Biology, Physics or Chemistry and Chemical Biology. Faculty associated with the Center for Systems Biology have access to facilities and opportunities for collaborative research through the Bauer Core facilities, the Center for Nanoscale Systems, the Broad Institute, and the Center for Brain Science.

Applications, including letters of reference, are due by Nov 15, 2012. Please submit a curriculum vitae, statement of research (≤5 pages, including a summary of previous research accomplishments), teaching statement and PDFs of ≤3 publications to http://academicpositions.harvard.edu/postings/4214. All files must be submitted electronically in PDF or Word format. During the application process you will be asked to give the e-mail addresses of at least three colleagues who have agreed to write letters of recommendation for you. Please allow at least 1 week for your referees to upload their letters.

*Applications from, or nominations of, women and minority candidates are encouraged. Harvard is an affirmative action/equal opportunity employer.*

NW227747R

---

Department of Health and Human Services
National Institutes of Health

■ ◆ ✶ ✸

National Institute on Aging

## Division of Geriatrics and Clinical Gerontology

**http://www.nia.nih.gov/about/offices/division-geriatrics-and-clinical-gerontology-dgcg**

### Health Scientist Administrator or Medical Officer

The National Institute on Aging's extramural Division of Geriatrics and Clinical Gerontology (DGCG) is seeking an individual to serve as a scientific administrator in its Clinical Gerontology Branch, to oversee and develop a program of grant-supported clinical and translational biomedical aging research.

Developmental aspects of the position include integration of findings from mechanistic, observational and intervention studies to identify research opportunities to advance experimental therapeutics for healthy aging and treatment of age-related conditions, and organizing initiatives to implement such research. Administrative aspects of the position include managing a portfolio of grants and grant applications on clinical and translational research on factors affecting aging over the life span and interventions influencing these factors.

Depending on training and qualifications, appointment to this position may be made either as a Health Scientist Administrator or as a Medical Officer. Candidates should have an M.D., Ph.D., or equivalent doctoral degree, and biomedical research experience. The ideal candidate will have training in a biomedical research field; experience in planning, design, administration, and analysis of clinical research; experience in scientific grant application writing or review; and demonstrated skills in administration and communication. Prior experience in experimental therapeutics and pharmaceutical R&D (e.g., identification of therapeutic targets, drug discovery, and knowledge of regulatory requirements) is highly desirable, but not required. Experience in relevant fields such as clinical physiology, gerontology, and genomics is also particularly valuable.

NIA's Division of Geriatrics and Clinical Gerontology (DGCG) is located in downtown Bethesda, Maryland. NIA is one of the National Institutes of Health (NIH), the premier biomedical research center for the Nation and the world. The 27 NIH Institutes and Centers employ approximately 18,000 persons in a wide array of jobs supporting efforts for improving health.

Salary is commensurate with experience and qualifications. If you have questions about this vacancy, please contact Caroline McCullough at caroline.mccullough@nih.gov or 410-558-8072. To apply and to review qualifications, evaluation criteria, application instructions, and salary and benefits information, please visit www.usajobs.gov and search for Health Scientist Administrator GS-601-13/14 announcement number NIH-NIA-DE-12-631912 (open to all US citizens) and/or Health Scientist Administrator GS-601-13/14 NIH-NIA-MP-12-631913 (open to current/former federal employees) and/or Health Scientist Administrator GS-601-15 announcement number NIH-NIA-DE-12-631897 (open to all US citizens) and/or Health Scientist Administrator GS-601-15 NIH-NIA-MP-12-631900 (open to current/former federal employees) and/or Medical Officer GS-602-13/14 announcement number NIH-NIA-DH-12-631924 (open to all US citizens) and/or Medical Officer GS-0602-15 announcement number NIH-NIA-DH-12-631925 (open to all US citizens). US citizenship is required. Applications will be accepted from September 11, 2012 to September 20, 2012.

*DHHS and NIH are Equal Opportunity Employers*

NW227662R

---

STANFORD
SCHOOL OF MEDICINE

*Stanford University Medical Center*

## Basic or Translational Vision Science Faculty Position Stanford University

The Stanford Institute for Neuro-Innovation & Translational Neurosciences (SINTN) and the Department of Ophthalmology seek to appoint a new faculty member in the area of basic or translational vision science. In addition to leading a research laboratory, the faculty member will be expected to mentor students and post-doctoral fellows. The appointment will be made at the Assistant, Associate or Full Professor level in the University Tenure Line, commensurate with the applicant's experience. The predominant criterion for appointment in the University Tenure Line is a major commitment to research and teaching.

**Please submit a curriculum vitae, statement of research interests and 3 reference letters via academicjobsonline.org by October 31, 2012 to:**

Brian Wandell, PhD
Isaac and Madeline Stein Family Chair
Director, Stanford's Center for Neurobiological Imaging
Chair, Search Committee
c/o Kristy Verhines
Stanford University School of Medicine
Lorry Lokey Stem Cell Research Building, G1078
265 Campus Drive
Stanford, CA 94305-5453

Stanford University is an equal opportunity employer and is committed to increasing the diversity of its faculty. It welcomes nominations of and applications from women and members of minority groups, as well as others who would bring additional dimensions to the university's research and teaching missions.

NW227918R

Memorial Sloan-Kettering Cancer Center

*The Best Cancer Care. Anywhere.*

# FACULTY POSITION

## Cell Biology Program

The Cell Biology Program, Sloan-Kettering Institute (www.ski.edu) has initiated a search for tenure-track faculty members. We are interested in outstanding individuals who have the potential to develop an innovative, independent research program that complements and enhances our existing strengths.

Candidates with research interests in exciting areas of eukaryotic cell biology, including aspects of stem cell biology, and using a variety of experimental approaches and systems are encouraged to apply.

New faculty will be eligible to hold graduate school appointments in the Gerstner Sloan-Kettering Graduate School of Biomedical Sciences, the Weill Cornell Graduate School of Medical Sciences of Cornell University, as well as the Tri-Institutional MD/PhD Training Program. Sloan-Kettering has an outstanding infrastructure as well as state-of-the-art core facilities resources, and we are expanding our research programs.

The deadline for applications is **November 1, 2012**. Interested candidates should visit **http://facultysearch.ski.edu** to access the on-line faculty application. Please visit the site as soon as possible, as it contains important information regarding the required application materials, including deadlines for submission of letters of reference. Informal inquiries may be sent to **Tiffany Lennon** at **lennont@mskcc.org** or **Dr. Alan Hall, Chair, Cell Biology Program** at **halla@mskcc.org**.

**facultysearch.ski.edu**

MSKCC is an equal opportunity and affirmative action employer committed to diversity and inclusion in all aspects of recruiting and employment. All qualified individuals are encouraged to apply.

NW228223R

---

## FACULTY POSITION IN BIOCHEMISTRY

Applications for a tenure-line faculty position (preferably at the Assistant or Associate Professor level) applying NMR to biological problems are invited by the Department of Biochemistry and School of Medicine at the University of Missouri in Columbia. We encourage individuals applying NMR to biological questions in many areas including, but not limited to metabolomics of disease, membrane proteins, intrinsically disordered proteins, protein complexes, solids and tissues, and biophysical enzymology. The Bruker 800, 500, and Varian 600 MHz spectrometers available on campus are each equipped with a cryogenic probe. Also available on campus are Bruker Avance III 7T micro-MRI and Siemens Trio 3T MRI systems capable of in vivo spectroscopy. The University is noted for interdisciplinary research in life sciences, including biophysical areas. Position qualifications include a Ph.D. or M.D. in biochemistry, biophysics or a related field, and postdoctoral experience. The successful applicant will develop or continue an outstanding, independently funded research program and contribute to departmental teaching activities.

Submit the following information to http://hrs.missouri.edu/find-a-job/academic/index.php

1) a cover letter;
2) curriculum vitae;
3) a 4-5 page description of current and planned research; and
4) the names and contact information of three references. Please use the Job Opening Number **8537**.

Application review will begin November 1st, 2012.

MU is an EEO/AA/ADA Employer, and encourages applications from women and minorities. For ADA accommodations, contact our Human Resource Services (573) 882-7976.

NW228163R

---

MGH 1811

HSCI HARVARD STEM CELL INSTITUTE

VE RI TAS

**The Center for Regenerative Medicine (CRM) at Massachusetts General Hospital and the Harvard Stem Cell Institute** invites applications for a tenure-track Assistant Professor position. Outstanding scientists in the field of vertebrate stem cell or regenerative biology will be considered. Candidates with a research program focused on tissue regeneration, plasticity and function and who have a demonstrated interest in application to human disease are especially welcome. Successful candidate(s) will be members of the **Harvard Stem Cell Institute** and faculty of **Harvard University**. Candidates must hold a PhD and/or MD and have a history of innovative, interactive research. Women and minority candidates are urged to apply. MGH is an Equal Opportunity/Affirmative Action Employer.

Applicants should send an electronic copy of a 1) letter of interest, 2) research plan, 3) current curriculum vitae and 4) three letters of recommendation to Dr. David Scadden c/o Miklosik.Mona@mgh.harvard.edu, Center for Regenerative Medicine Search Committee, Massachusetts General Hospital, 185 Cambridge St., CPZN 4258, Boston, MA 02114.

NW227622R

---

### Yale School of Forestry & Environmental Studies

***Junior or Senior Faculty Position in Tropical Forest Management***

The Yale School of Forestry & Environmental Studies has an open rank, tenure-track opening for an exceptional natural scientist whose research focuses on the conservation and management of tropical forest resources. Ideal candidates will conduct interdisciplinary, field-based, hypothesis-driven research that bridges the social and biophysical sciences. They will also demonstrate potential for collaborating with natural, physical and social scientists at Yale and within a strongly interdisciplinary School.

Areas of research could include: the science, policy and management of forest resources; biodiversity; climate-change related interventions; and the intersection between forest resource management, agriculture and land use. The ability to develop student field experiences, teach classes, and conduct research in both temperate and tropical systems would be an asset.

The successful candidate will be expected to develop an internationally recognized research program, and will teach introductory and advanced classes in the Yale School of Forestry & Environmental Studies and also in Yale College.

Applicants should submit a curriculum vitae, a statement of research and teaching interests, and the names of four references via email with the subject line "Environment Faculty Searches – Tropical Forest Management" to fesdeansoffice@yale.edu or via surface mail to:

Tropical Forest Management Faculty Search
c/o Angela Kuhne, Dean's Office
Yale School of Forestry & Environmental Studies
195 Prospect Street
New Haven, CT 06511
USA

Prior to applying, candidates should explore the School's website (www.environment.yale.edu) and consider how their expertise can strengthen or complement the strengths of the existing faculty of the School.

Applications received by October 5, 2012 will receive full consideration. For more information about the position, contact Assistant Dean Angela Kuhne at angela.kuhne@yale.edu.

***Yale University is an affirmative action/equal opportunity employer. Men and women of diverse racial/ethnic backgrounds and cultures are encouraged to apply***

NW227529RL

# Professor Positions of various levels in Crop Biology Research Center of China

To implement the '2011 Plan' launched by the Ministry of Education and Ministry of Finance of China, China Agricultural University together with North West Agriculture and Forestry University and Huazhong Agricultural University, has establish a Coordinated Crop Biology Research Center (CBRC). CBRC integrates the main resources and research strength of maize, rice and wheat in China, aiming to become one of the world's leading research center for crop sciences

Highly motivated and talented candidates (all nationalities) are encouraged to apply for faculty positions at various levels, including senior and junior principle investigators (PIs) and research associates. All PIs will be provided with competitive packages including set-up and annual operating funds, and attractive salaries comparable to the international standard. To view the complete position description, qualifications, and application instructions, please visit http://2011.cau.edu.cn/art/2012/8/16/art_6511_137035.html, or feel free to contact the search committee office of CBRC by telephone (86-10-62734762) or email (2011@cau.edu.cn). Review of application will start immediately until the positions filled.

JP228227R

---

# Penn Medicine

## Assistant Professor (Tenure Track)

The Department of Radiation Oncology, Division of Radiation Research at the Perelman School of Medicine, University of Pennsylvania, invites applications for a tenure-track faculty position at the Assistant Professor level. Candidates with a PhD or an MD/PhD degree, strong publication records in any area of Tumor or Radiation Biology and particularly those with research interest in normal tissue radiobiology, tumor microenvironment or tumor stem cells, are encouraged to apply.

The Division of Radiation Research is comprised of 11 faculty engaged in innovative, translational research with a focus on the study of the tumor microenvironment, radiation response modifiers, photodynamic therapy and proton therapy. Research in the Division is supported by multiple grants from the NIH, DoD, NASA, private foundations and industry. Faculty in the Division can have cross-appointments to other Departments and can belong to different graduate groups such as Cancer Biology and Pharmacology.

The Division has recently moved to a brand new, contiguous laboratory space in the Translational Research Building, in very close proximity to the clinic. State-of-the-art shared instrumentation includes a 3-D conformal irradiator for small animals, live cell microscopy, hypoxia chambers, multiple PDT lasers, etc. The TRC building has its own barrier and non-barrier animal facility and small animal imaging facilities which include optical, MR and CT imaging. A competitive start-up package and laboratory space will be provided.

For more information about the Department of Radiation Oncology, and the Division of Radiation Research please visit:
**http://www.xrt.upenn.edu/index.shtml**

Interested candidates should send a curriculum vitae, a one-page summary of their future research goals and the names of three references, preferably in a single PDF, to **Ms. Rebecca Wilson**, at: **Rebecca.Wilson@uphs.upenn.edu**

The deadline for applications is December 1, 2012.

We seek candidates who embrace and reflect diversity in the broadest sense.

*The University of Pennsylvania is an equal opportunity, affirmative action employer.*

NW228129RL

---

## Neurobiology of Disease Faculty Position

### **Gladstone Institutes** and the University of California, San Francisco

The Gladstone Institute of Neurological Disease and the University of California, San Francisco (UCSF) invite applications for a faculty position at the level of Assistant, Associate, or full Professor. Primary criteria for appointment will be outstanding records of innovative research and academic performance, including landmark papers in leading journals, as well as proven abilities or high potential to establish a rigorous and substantial independent research program, provide exemplary mentorship, and engage in fruitful collaborations.

The successful candidate will join an interactive group of investigators in Gladstone's state-of-the-art research facility at UCSF's Mission Bay campus. S/he will have joint appointments in the Gladstone Institute of Neurological Disease and in the Neuroscience Graduate Program and the Biomedical Sciences Graduate Program at UCSF. Excellent salary support is provided. Women and underrepresented minorities are especially encouraged to apply. The search will continue until the position is filled. However, to ensure full consideration, applications should be received by **November 1, 2012**.

Please send curriculum vitae, description of achievements and research interests, and the names of three references to:

**GIND/UCSF Search Committee**
**1650 Owens Street**
**San Francisco, CA 94158**
**gindsearch@gladstone.ucsf.edu**

**GLADSTONE**

The Gladstone Institutes and UCSF are Affirmative Action/Equal Opportunity Employers. They undertake affirmative action to assure equal employment opportunity for underutilized minorities and women, for persons with disabilities, and for Vietnam-era veterans and special disabled veterans.

NW228178R

---

# UCDAVIS
UNIVERSITY OF CALIFORNIA

## ASSISTANT & ASSOCIATE PROFESSORS (3 positions)

### COLLEGE OF BIOLOGICAL SCIENCES UNIVERSITY OF CALIFORNIA, DAVIS

The College of Biological Sciences, University of California (www.biosci.ucdavis.edu), Davis, invites applications for three positions at the levels of Assistant (tenure-track) or Associate (tenured) Professor. Successful candidates will join a world-class research and academic environment in one of only two dedicated colleges of biology in the U.S. The unique collaborative and interdisciplinary research environment in the College of Biological Sciences comprises 120 faculty organized into five Departments (Evolution and Ecology; Microbiology; Molecular and Cellular Biology; Neurobiology, Physiology and Behavior; and Plant Biology) and three interdisciplinary Centers (Genome; Neuroscience; and Population Biology) that span multiple academic units. This is a broad search to identify outstanding candidates investigating fundamental questions in modern biology using any experimental system from microbes to plants to humans. We seek to build on our programs in structural and molecular biology, biophysics and biochemistry, cellular and developmental biology, physiology, genetics, and systems biology. Candidates working in any related area are encouraged to apply. Appointments to the most appropriate academic units will be determined in consultation with the successful candidates.

Candidates must have an outstanding record of research achievement and will be expected to develop a strong research program investigating fundamental questions in modern biology. Successful candidates will also participate in normal teaching activities. A Ph.D. (or equivalent) is required. Applicants should submit materials online at: https://recruit.ucdavis.edu/. Please include a curriculum vitae, descriptions of current research (up to 2 pages) and future research program (up to 3 pages), a one page summary of teaching interests and experience, and up to three publications. Applicants should also arrange to have three referees submit supporting letters online at the above website. Closing date: Open until filled, but all application materials, including letters of recommendation, must be received by October 15, 2012 to be assured full consideration. Administrative contact: Sally DiVecchia (svdivecchia@ucdavis.edu). Faculty contact: Stephen Kowalczykowski (via Carol Nickles: cnickles@ucdavis.edu).

*The University of California is an Equal Opportunity/Affirmative Action Employer with a strong institutional commitment to the development of a climate that supports equality of opportunity and respect for differences.*

NW227743R

## TENURE-TRACK ASSISTANT PROFESSORSHIP IN CHEMISTRY

**Harvard University**
**Faculty of Arts and Sciences**
**Department of Chemistry**
**& Chemical Biology**
**Cambridge, MA**

Candidates are invited to apply for a tenure-track assistant professorship in physical chemistry. The appointment is expected to begin on July 1, 2013. The tenure-track professor will be responsible for teaching at the undergraduate and graduate levels. We are seeking candidates who have an outstanding research record and a strong commitment to undergraduate and graduate teaching. Ph.D. required by expected start date. Candidates should arrange to have three letters of recommendation sent independently and provide a curriculum vitae, statement of teaching philosophy, list of publications, and outline of their future research plans. All applications and supporting materials must be submitted via the ARIeS portal (https://academicpositions.harvard.edu/postings/4247) no later than October 15, 2012.

*Harvard is an Equal Opportunity/Affirmative Action employer. Applications from women and minorities are strongly encouraged.*

NW227956R

---

## TENURE-TRACK ASSISTANT PROFESSORSHIP IN CHEMISTRY

**Harvard University**
**Faculty of Arts and Sciences**
**Department of Chemistry**
**& Chemical Biology**
**Cambridge, MA**

Candidates are invited to apply for a tenure-track assistant professorship in inorganic chemistry, broadly defined to include reaction chemistry, with potential connections to energy-related research. The appointment is expected to begin on July 1, 2013. The tenure-track professor will be responsible for teaching at the undergraduate and graduate levels. We are seeking candidates who have an outstanding research record and a strong commitment to undergraduate and graduate teaching. Ph.D. required by expected start date. Candidates should arrange to have three letters of recommendation sent independently and provide a curriculum vitae, statement of teaching philosophy, list of publications, and outline of their future research plans. All applications and supporting materials must be submitted via the ARIeS portal (https://academicpositions.harvard.edu/postings/4246) no later than October 15, 2012.

*Harvard is an Equal Opportunity/Affirmative Action employer. Applications from women and minorities are strongly encouraged.*

NW227955R

---

## Fellowships

The Radcliffe Institute for Advanced Study at Harvard University annually awards academic-year fellowships that enable scientists and mathematicians to pursue innovative research while participating in the Institute's diverse scholarly community.

While at Radcliffe, these fellows are able to work in Harvard and Boston-area labs and with faculty and other fellows to explore new avenues in their research.

Scientists in any field who have a doctorate in the area of the proposed project (by December 2011) and at least one published article or monograph are eligible to apply for a Radcliffe Institute fellowship. The stipend amount of $70,000 is meant to complement sabbatical leave salaries of faculty members. Residence in the Boston area is required, as is participation in the Institute community.

Applications for 2013–2014 are due by November 1, 2012.

For more information, please visit **www.radcliffe.harvard.edu.**

RADCLIFFE INSTITUTE
FOR ADVANCED STUDY
HARVARD UNIVERSITY

NW227921R

---

## ASSISTANT PROFESSOR POSITIONS AVAILABLE AT PRINCETON UNIVERSITY

The Lewis-Sigler Institute for Integrative Genomics at Princeton University invites applications for a tenure-track faculty position at the Assistant Professor level; faculty positions at the Lewis-Sigler Institute are joint appointments with an appropriate home department. We are seeking outstanding scientists with strong interest and experience in quantitative, systems-level approaches to understanding any area in modern experimental biology. Candidates with expertise in experimental methods for high-throughput biology are particularly encouraged to apply. A strong record of experimental work and quantitative analysis is essential. The successful candidate will have research laboratories at the Institute, and teaching responsibilities (both graduate and undergraduate) will be shared between the Institute and the home department.

The Lewis-Sigler Institute for Integrative Genomics, housed in the Carl Icahn Laboratory at Princeton University, was established to innovate in research and teaching at the interface of modern biology and the more quantitative sciences. The Institute has invested heavily in providing state-of-the-art infrastructure, including facilities for High Throughput Sequencing and Microarrays, Advanced Imaging, Mass Spectrometry, and Computation.

**Essential Qualifications**

All applicants must have a Ph.D., M.D., or equivalent degree. In addition, applicants must have a very strong record of research productivity, demonstrate the ability to develop a rigorous research program, and be committed to teaching at both the undergraduate and graduate levels.

**How to Apply**

Applications must be submitted online at: http://jobs.princeton.edu/applicants/Central?quickFind=62744 and should include a cover letter, curriculum vitae, a two-page research description, as well as contact information for three references. Applications will be reviewed beginning on October 15, 2012, and applications arriving after November 15, 2012 may not be considered.

*Princeton University is an equal opportunity employer and complies with applicable EEO and affirmative action regulations.*

NW227740RL

---

## Yale University - **Chemical Biology Institute**

The Chemical Biology Institute, a cornerstone of Yale's new West Campus research enterprise, invites applications for tenure track positions at the **ASSISTANT PROFESSOR** level to commence 1 July 2013. Faculty associated with this Institute will hold primary appointments in any of several life science or physical science departments within the Faculty of Arts and Sciences, the School of Engineering and Applied Science, or the Yale School of Medicine. **At this time, we seek candidates whose research interests and accomplishments align best with the Institute and Yale's School of Engineering and Applied Science** (http://www.seas.yale.edu).

Successful candidates will be creative teacher-scholars with reputations for outstanding research at the combined interface of engineering, chemistry, biology and/or medicine, and must possess a Ph.D. in a relevant discipline, and must be committed to excellent pedagogy in engineering education.

Applicants should create a profile at **https://academicjobsonline.org/ajo/jobs/1725** and upload a statement of research plans, curriculum vitae, and up to five reprints of published work(s). Applicants should also arrange for three references to upload their letters of recommendation.

For further information, contact Jessica Sorensen at jessica.sorensen@yale.edu or PO Box 27392, West Haven, CT 06516-7392.

The review of applications will begin on **15 October 2012** and proceed until suitable candidates are identified.

Yale University is an affirmative action, equal opportunity employer. Yale values diversity among its faculty, students, and staff and strongly encourages applications from women and underrepresented minorities.

NW227922RL

---

## Princeton University
### Department of Chemistry
### Assistant Professor - All Areas

The Department of Chemistry at Princeton University invites applications for a tenure-track assistant professor position in all areas of chemistry. Candidates should have a strong commitment to research and to teaching at the undergraduate and graduate levels, and are expected to have completed the Ph.D. in chemistry or a related field at the time of appointment.

Applicants should submit a description of research interests, curriculum vitae, a list of publications, and contact information for three referees online at http://jobs.princeton.edu/appli cants/Central?quickFind=62756.

The search committee will begin review of applications on 10/17/2012 and will continue until the position is filled.

*Princeton University is an Equal Opportunity employer and complies with applicable EEO and affirmative action regulations.*

NW227919RL

Imperial College
London

# RECRUITING NOW

## to a wide range of posts including...

MRC | Clinical Sciences Centre

The Clinical Sciences Centre is a leading MRC Research Institute at the Hammersmith Campus of Imperial College, London, conducting wide-ranging programmes in basic, applied and clinical research. Our aim is to train and mentor the next generation of scientists, to carry out world class research, and to facilitate the translation of this understanding into improved diagnosis and treatment.

### Post-doctoral and research support positions:

Investigating the significance of brain-gut interactions in the regulation of food intake, nutrient storage and metabolic homeostasis using a combination of molecular, metabolic and behavioural methods in the fruit fly Drosophila melanogaster.
**Ref: IRC63368 / HM2012133**

Studying the mechanisms of long-range transcriptional regulation in embryonic development, and their implications for genome evolution and disease – a background in computational comparative genomics and/or analysis of next-generation sequencing-based experiments and knowledge of molecular biology are required.
**Ref: IRC63628 / IRC63629**

Investigation of novel molecular and epigenetic mechanisms of dedifferentiation, reprogramming and plasticity in adult tissue regeneration.
**Ref: IRC63366**

Determining the impact of cell cycle control for pluripotent self-renewal and for dominant reprogramming, and to identify factors that enhance conversion of somatic cells towards pluripotency.
**Ref: HM2012143/145**

Experimental and computational analysis of gene regulation by transcription factors eg Ikaros, CTCF and chromatin proteins eg cohesin, condensin.
**Ref: HM2012092**

A **Masters/PhD level** informatics position for 9 months in Cancer Genomics - experience with R, SQL and web development desirable.
**Ref: IRC63589**

To find out more go to
**www.imperial.ac.uk/employment**

**www.csc.mrc.ac.uk**

U228038R

---

Penn
UNIVERSITY *of* PENNSYLVANIA

## UNIVERSITY OF PENNSYLVANIA

The School of Arts and Sciences at the University of Pennsylvania invites applications for a tenure-track assistant professor appointment in evolution, broadly interpreted. We are interested in exceptional scientists and mathematicians who have well-developed research programs employing mathematical or computational techniques to study the evolution of dynamical processes far from equilibrium in the context of any of the following: biology, chemistry, or materials from the molecular to the systems scale, language, geology, psychology, or the environment. The successful candidate's primary appointment will be in a single department in the natural sciences: Biology, Chemistry, Earth and Environmental Science, Linguistics, Mathematics, Physics and Astronomy, or Psychology. Secondary appointments in other departments can be arranged, as appropriate. This appointment will be the first in a cluster of appointments across the natural sciences in various aspects of evolution; the successful candidate should therefore have a strong interest in building such a program and in interacting with researchers from other disciplines whose research lies within the overarching theme of evolution. The successful candidate will teach courses in his or her home department and will participate in the development of curriculum pertinent to the theme of the cluster.

Applications should be submitted on-line at **facultysearches.provost.upenn.edu/applicants/Central?quickFind=51089** and include a curriculum vitae, a research statement that includes the candidate's perspective on how she or he fits into one of the core departments, links to no more than three journal publications, and the contact information for three individuals who will be contacted by the University with instructions on how to submit a letter of recommendation. Review of applications will begin 16 November 2012 and will continue until the position is filled.

The University of Pennsylvania is an affirmative action/equal opportunity employer and is strongly committed to establishing a diverse faculty:
http://www.upenn.edu/almanac/volumes/v58/n02/diversityplan.html

NW228124RL

---

## FACULTY POSITION
## Princeton University

**The Molecular Biology Department at Princeton University invites applications for a senior tenured faculty position at the Full or Associate Professor level. We are seeking an outstanding investigator to address fundamental questions in all areas of Molecular Biology.**

The University has a strong commitment to interdisciplinary studies, especially in the areas of systems biology, imaging, genomics, biophysics and evolution. The department has high-level computing and microscope facilities, DNA array and high throughput sequencing technologies, mass spectrometry, and state of the art vivarium. Applicants must have an excellent record of research productivity and demonstrate the ability to develop a rigorous research program.

All applicants must have a Ph.D. or equivalent degree and a commitment to teaching at the undergraduate and graduate levels. Applications must be submitted online at http://jobs.princeton.edu, requisition **#1200537** and should include a cover letter, curriculum vitae, a two-page research description, and contact information for three references. All materials must be submitted as PDF files.

Screening of applications will begin
1 October, 2012.

*Princeton University is an equal opportunity employer and complies with applicable EEO and affirmative action regulations.*

NW228074R

---

## FACULTY POSITION
## Princeton University

**The Molecular Biology Department at Princeton University invites applications for a tenure-track faculty position at the Assistant Professor level. We are seeking an outstanding investigator using multicellular model systems to address fundamental questions in Multiscale Cell Dynamics.**

The University has a strong commitment to interdisciplinary studies, especially in the areas of systems biology, imaging, genomics, biophysics and neuroscience. The department has high-level computing and microscope facilities, DNA array and high throughput sequencing technologies, mass spectrometry, and state of the art vivarium. Applicants must have an excellent record of research productivity and demonstrate the ability to develop a rigorous research program.

All applicants must have a Ph.D. or equivalent degree and a commitment to teaching at the undergraduate and graduate levels. Applications must be submitted online at http://jobs.princeton.edu, requisition **#1200533** and should include a cover letter, curriculum vitae, a two-page research description, and contact information for three references. All materials must be submitted as PDF files.

Screening of applications will begin
1 October, 2012.

*Princeton University is an equal opportunity employer and complies with applicable EEO and affirmative action regulations.*

NW228075R

LMU
LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

GRADUATE SCHOOL OF
QUANTITATIVE BIOSCIENCES MUNICH

QB

## Administrative coordinator

The newly established **Graduate School of Quantitative Biosciences Munich (QBM)** seeks an administrative coordinator to manage its day-to-day operation. The School is designed to prepare young life scientists for the emerging era of quantitative, systems-oriented bioscience and brings together a range of disciplines, including biochemistry, medicine, bioinformatics, experimental and theoretical (bio-)physics, and applied mathematics. It offers an innovative, international PhD program that combines interdisciplinary research with an intensive interdisciplinary course curriculum, with the aim of enabling students to work effectively at the interface between traditional disciplinary boundaries. As the administrative coordinator, you will be responsible for all aspects of the daily operation and administration of the School. You will be assisted by an administrative secretary as well as by two lecturers and work closely with the speaker and scientific board of the School.

QBM is funded through the German Excellence Initiative and embedded in a strong and varied research environment in the biosciences in Munich. The School is based both at the central campus of the Ludwig-Maximilians-Universität (LMU) in downtown Munich and at the life sciences campus in Munich-Grosshadern/Martinsried.

**Your tasks**

- General administration of the School, including finances and personnel,
- managing student recruitment (advertisement, review of applications and selection for interviews, organizing the interview week),
- planning and organizing events (lectures, meetings),
- assisting students with administrative tasks,
- interfacing with academic institutions and principal investigators,
- representing the School to the scientific community and public.

**Your profile**

- PhD in natural sciences or related field,
- postdoctoral and administrative experience,
- excellent organizational, interpersonal and team skills,
- international experience and fluent communication in both English and German.

The salary is according to the TV-L (E 14). The position is available immediately and will initially be for two years, but with the possibility of extension. The LMU is an equal opportunity employer. Disabled candidates with essentially equal qualifications will be given preference.

Please send a single pdf-file with your application (including cover letter, CV, degrees, references) by **September 30, 2012** to: **Prof. Dr. Ulrike Gaul (speaker of QBM), email: qbm-office@genzentrum.lmu.de,** http://www.qbm.lmu.de.

W227816R

---

Memorial Sloan-Kettering
Cancer Center
*The Best Cancer Care. Anywhere.*

## Faculty Positions in Cancer Biology

Applications are invited for tenure-track faculty positions in the Cancer Biology and Genetics Program of the Sloan-Kettering Institute, Memorial Sloan-Kettering Cancer Center (www.ski.edu). Successful candidates will carry out independent research on the genesis, progression, prognosis, prevention and treatment of cancer that synergizes with ongoing efforts at the Center. Areas of special interest include, but are not limited to: cancer genetics, cancer stem cells, metastasis, tumor microenvironment, inflammation and cancer, and animal models of cancer.

New faculty members will join an interactive, interdisciplinary community of scientists and clinicians at the Center, which offers an outstanding basic and translational research environment within expanded state-of-the-art research facilities. Faculty will be eligible to hold graduate school appointments in the Gerstner Sloan-Kettering Graduate School of Biomedical Sciences, the Weill Cornell Graduate School of Medical Sciences of Cornell University, as well as the Tri-Institutional MD/PhD Training Program.

### Cancer Biology & Genetics Faculty

**Robert Benezra, PhD** – Chromosome Instability and Cancer
**Eric Holland, MD/PhD** – Brain Tumorigenesis
**Kitai Kim, PhD** – Induced Pluripotent Stem Cells
**Robert Klein, PhD** – Cancer Genetics
**Johanna Joyce, PhD** - Tumor Microenvironment
**Scott Lowe, PhD** - Tumor Suppressor Genes
**Joan Massague, PhD (Chairman)** - Cell Signaling and Metastasis
**Christine Mayr, MD, PhD** - mRNA Control Mechanisms
**Kenneth Offit, MD** – Cancer Genetics
**Craig Thompson, MD** - Cancer Metabolism
**Andrea Ventura, MD, PhD** – microRNAs in Development and Cancer
**Hans-Guido Wendel, MD** – Genetic Basis for Drug Resistance
**Richard White, MD, PhD** - Modeling Metastasis in Zebrafish

The deadline for applications is November 1, 2012. Interested candidates should visit **http://facultysearch.ski.edu** to access the on-line faculty application. Please visit the site as soon as possible, as it contains important information on the required application materials, including deadlines for submission of letters of reference. Inquiries may be sent to Gregory Erskine at **erskineg@mskcc.org**.

**facultysearch.ski.edu**

MSKCC is an equal opportunity and affirmative action employer committed to diversity and inclusion in all aspects of recruiting and employment. All qualified individuals are encouraged to apply.

NW227654R

---

# nature

## Cross-Title Associate Editor

Nature Publishing Group (NPG) is a division of Macmillan Publishers Ltd, dedicated to serving the academic, professional scientific and medical communities. For further information please visit www.nature.com

Nature Reviews, Nature Publishing Group's noted life science series, includes eight clinical titles that publish high-quality, authoritative review and opinion articles http://www.nature.com/reviews/index.html. We now have an exciting opportunity for a permanent, full-time Associate Editor who will work across all eight clinical Nature Reviews titles.

The Associate Editor's role is busy and varied. You will work closely with the in-house journal teams, and liaise directly with their internationally renowned Advisory Boards and contributors.

Key tasks include commissioning and editing for all editorial sections, organizing peer review, and writing Research Highlights.

The job is based in our modern London offices but some travel to conferences may be required.

Preference will be given to candidates with the following key attributes: a clinical or biological science degree to PhD level (in an area relevant to at least one of the journals' subject areas); a demonstrable ability to edit text for sense, clarity, factual correctness and language; experience that indicates a proactive approach; and excellent organizational skills. Training will be provided, but prior editorial experience will be advantageous.

To apply, please submit your CV and a covering letter confirming your current salary and quoting reference **NPG/095/12** via our online recruitment site: **http://jobs.macmillan.com**

All applicants must be able to demonstrate the right to live and work in the UK in order to be considered for the vacancy.

**Closing Date: Wednesday 12th September 2012**

nature publishing group npg

IN228200R

---

Bringing jobseekers and recruiters together.

**Academic • Industry • Government**

Search now at **naturejobs.com**

naturejobs
making science work

**Imperial College London**

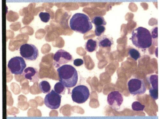
**Imperial College Healthcare NHS**
NHS Trust

# Major initiative in an internationally renowned centre of excellence for Haematology

Imperial College London is making a major investment in Haematology, a unit with an outstanding international reputation. This was the first academic Department of Haematology in the UK, founded in 1966 and since that time has consistently contributed to advances in all aspects of malignant, non-malignant and laboratory haematology. We plan to make a number of senior academic appointments in line with existing preclinical and clinical strengths and to develop a world-leading portfolio of translational research in collaboration with Imperial College Healthcare NHS Trust. Existing well-funded research within the Centre for Haematology focuses on cellular and immunotherapy, the biology and management of chronic myeloid leukaemia and myeloma, haemostasis and thrombosis, and congenital and acquired disorders of Haemopoiesis. Haematology is one of the core themes of the NIHR Biomedical Research Centre and has an active clinical trials programme with both commercial and academic initiated clinical studies.

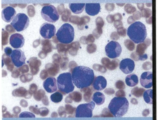
Imperial College London provides an outstanding multidisciplinary environment for basic, clinical and translational research. The Centre has substantial research facilities including recently constructed laboratory space at the Hammersmith Hospital and an outstanding strong patient base (with significant research network support), a JACIE, HTA and MHRA accredited cellular therapy programme, embedded research tissue biobanks, excellent links with state-of-the-art molecular pathology and a total commitment to translational medicine.

Imperial College
London
← Imperial Centre for
Translational &
Experimental Medicine

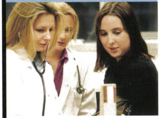


We are especially interested to hear from both non-clinical and clinical academics with highly innovative research objectives and a track record of high quality publications. We seek to develop a multi-disciplinary Faculty with interfaces between:

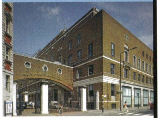
- Stem cell biology
- Stem cell transplantation
- Molecular pathology
- Epigenetics and genomics of haematological malignancy
- Haemostasis and thrombosis



The level of appointment (Professor, Reader, Senior Lecturer or Lecturer) and the generous supporting package will be determined according to experience. There is provision for individually negotiated appointments and appropriately tailored job descriptions to encourage innovative research approaches or translational work.



Imperial College London provides an outstanding multidisciplinary environment and this is a unique career opportunity for individuals with the vision, leadership ability and ambition to be among the best in the field. Interested individuals are encouraged to make informal enquiries to the Head of Centre, Prof. Jane Apperley (j.apperley@imperial.ac.uk). Applications including a letter describing research and career objectives and a full curriculum vitae should be sent to Alice Ashley-Smith on or before 30 September 2012 (Tel: 0208 383 2152; email: a.ashley-smith@imperial.ac.uk).



*Committed to equality and valuing diversity. We are also an Athena Bronze SWAN Award winner, a Stonewall Diversity Champion and a Two Ticks Employer.*

POSITIVE ABOUT DISABLED PEOPLE

Stonewall
DIVERSITY CHAMPION

U228239R

a**nature**conference

*Nature Immunology* and HudsonAlpha Institute for Biotechnology present:

# HudsonAlpha - Nature Conference on Immunogenomics

The first HudsonAlpha - Nature Conference on Immunogenomics will bring together four areas to shape this emerging field: scientists in genomics and genetics, immunology, bioinformatics and methodology, and clinical research. We expect our invited talks to cover a range of topics: large-scale immune sequencing projects of the repertoire of T- and B-cell diversity, and clinical implications; epigenetics and the immune system; developmental lineage of immune cells; genetic and environmental contributions to autoimmunity; drug development for autoimmune conditions; pharmacogenomics and the immune response; gene regulation and cell-type identity and high-resolution transcriptome analyses.

## October 1-3, 2012
## The Jackson Center, Huntsville, AL, USA

### KEYNOTE SPEAKERS

**Daniel Kastner** (National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, USA)
**Stephen Quake** (Stanford University, USA)
**Matthias Merkenschlager** (MRC Clinical Sciences Centre, UK)
**Ugur Sahin** (Translational Oncology at the University Medical Center of the Johannes Gutenberg University Mainz, Germany)

### SPEAKERS

**Christophe Benoist**
**S. Louis Bridges, Jr.**
**Christopher Glass**
**Jian Han**
**Robert Holt**
**John O'Shea**
**Robert Plenge**
**Harlan Robins**
**Ellen Rothenberg**
**Harinder Singh**
**Casey Weaver**
**Wenzhong Xiao**

### ORGANIZERS:

**S. Louis Bridges, Jr** (University of Alabama Birmingham, USA)
**Laurie Dempsey** (*Nature Immunology*, USA)
**Chris Gunter** (HudsonAlpha Institute for Biotechnology, USA)
**Jian Han** (HudsonAlpha Institute for Biotechnology / iCubate / iRepertoire, USA)
**Robert Holt** (Michael Smith Genome Sciences Centre, BC Cancer Agency, Canada)
**Robert Plenge** (Harvard Medical School / Brigham & Women's Hospital, USA)
**Ellen Rothenberg** (California Institute of Technology, USA)

**Abstract submission deadline EXTENDED to July 27, 2012**
**Early registration deadline: August 3, 2012**
For more information and to register visit:
**www.nature.com/natureconferences/haig2012**

nature
immunology

HUDSONALPHA
INSTITUTE FOR BIOTECHNOLOGY

IN225038E

10th
2002
2012
Anniversary

CRG
Centre
for Genomic
Regulation

## Anniversary Symposium

# 10 Years of Research, Education & Training at the CRG

## 18 October 2012

On-line registration at www.crg.eu/2012_symposium
Registration deadline: September 30, 2012

**Invited Speakers**

- **Angus LAMOND,** Wellcome Trust Centre for Gene Regulation and Expression, Dundee, United Kingdom
- **Tom MANIATIS,** Columbia University, New York, USA
- **Randy SCHEKMAN,** Howard Hughes Medical Institute, University of California at Berkeley, USA
- **Richard SCHELLER,** Genentech Research, San Francisco, USA
- **Austin SMITH,** Wellcome Trust Centre for Stem Cell Research, University of Cambridge, Cambridge, UK
- **Veronica VAN HEYNINGEN,** MRC Human Genetics Unit, Edinburgh, United Kingdom
- **Ada E. YONATH,** Weizmann Institute of Science, Rehovot, Israel

**Invited Alumni**

- **Marcus BUSCHBECK,** Institute of Predictive and Personalized Medicine of Cancer (IMPPC), Badalona, Spain
- **Marc GÜELL,** Harvard Medical School, Boston, USA
- **Albert JORDAN,** Molecular Biology Institute of Barcelona (IBMB-CSIC), Barcelona, Spain
- **Raul MÉNDEZ,** Institute for Research in Biomedicine (IRB Barcelona), Barcelona, Spain
- **Julia VON BLUME,** Max Planck Institute of Biochemistry, Martinsried, Germany

Members of the Board of Trustees:

Generalitat de Catalunya

UNIVERSITAT POMPEU FABRA

Sponsored by:

NOVARTIS SANOFI illumina Roche AB SCIEX life technologies

Centre for Genomic Regulation (CRG), PRBB Building, Dr. Aiguader 88, 08003 Barcelona (Spain) – Ph: +34 93 316 01 00 – www.crg.eu

W225526EL

---

# Stay in touch with the latest scientific news and research, anytime, anywhere

**With the new nature.com mobile app you can read scholarly content quickly and comfortably wherever you are**

Download now at
**nature.com/mobileapps**

Available on the App Store

Browse updated news and research each time you access

Read full text articles immediately, or save them for later

9:42 AM
nature.com
Login
Register
Back
neuroscience schizophrenia
Save
Results for: neuroscience schizo...
Recent articles
Saved
Search
OK

Please log in with your nature.com account. If you don't already have an account, tap 'Register' to create one - it takes less than two minutes.

NEWS AND VIEWS
Neuroscience: Excitatory view of a receptor
Lonnie P. Wollmuth & Stephen F. Traynelis

Neuroscience: Excitatory view of a receptor
Nature
one month ago

Behaviour: Flies on film
Nature

Saved
The current search terms will now appear in your list of saved searches.

2 months ago

Neuroscience in the era of functional genomics and

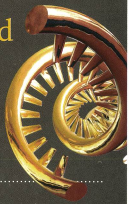
Set up saved searches on PubMed or nature.com

Access high resolution figures and zoom in to see fine detail

iPhone is a trademark of Apple Inc.

nature publishing group npg

*Wellcome Trust Genome Campus*



# Advanced Courses and *Scientific* Conferences

## Scientific Conferences

**Rat Genomics and Models**
3–6 December 2012

**Biomarkers for Brain Disorders**
3–5 February 2013

**Genomic Disorders: The Genomics of Rare Diseases**
10–13 April 2013

**Managing Mouse Colonies**
8–10 May 2013

**Applied Bioinformatics and Public Health Microbiology**
15–17 May 2013

**Exploring Human Host-Microbiome Interactions in Health and Disease**
8–10 July 2013

**Summer School on Biology of Social Cognition**
14–20 July 2013

**The Leena Peltonen School of Human Genomics**
18–22 August 2013

**The Genomics of Common Diseases**
7–10 September 2013 (Oxford, UK)

**Functional Genomics and Systems Biology**
21–23 November 2013

## Advanced Courses and Workshops

**Genomics and Clinical Microbiology**
20–25 January 2013

**Complex Disease Genetics for Clinicians**
24–25 January 2013

**Mathematical Models for Infectious Disease Dynamics**
4–15 February 2013

**Malaria Experimental Genetics**
3–9 March 2013

**Computational Molecular Evolution**
29 April–10 May 2013

**Molecular Approaches to Clinical Microbiology in Africa (Blantyre, Malawi)**
25 May–1 June 2013

**In silico Systems Biology**
25–29 June 2013

**Genetic Analysis of Multifactorial Diseases**
24–30 July 2013

**Exome Sequencing**
8–16 September 2013

**www.wellcome.ac.uk/hinxton**

**wellcome**trust

W227264EL

# naturejobs

professional
interesting
public
now
work
start
describe
read
work
like
comments
personal
write
know
identity
affect
changes
use
thoughts
post
others
blog
try
offline
life
online
find
make
traffic
see
changes
work
interesting
write
online
read

**How can you make more informed career or hiring decisions?**

Follow the blog on naturejobs.com to keep up-to-date with the latest scientific employment news from around the world.

**What do you think about the news stories affecting scientific careers?**

Start sharing your opinions today on the *Naturejobs* blog.

**www.naturejobs.com**

**REPROGRAPHICS: CC Media Group**, London EC1V 8AR UK and **The Charlesworth Group,** Wakefield, UK.
**PRINTED BY: William Gibbons & Sons Ltd,** UK; **Publishers Press**, Lebanon Junction, Ky, USA and **Obun Printing Co. Inc,** Tokyo, Japan

# Cold Spring Harbor Laboratory Meetings & Courses

## 2012 Fall / Winter Meetings

**Axon Guidance, Synapse Formation & Regeneration**
September 18 - 22
Alain Chedotal, Graeme Davis, Carol Mason, Yishi Jin

**Dynamic Organization of Nuclear Function**
September 27 - October 1
Edith Heard, Martin Hetzer, David L. Spector

**Germ Cells**
October 2 - 6
Tim Schedl, Erika Matunis, Laurinda Jaffe

**Molecular Genetics of Aging**
October 9 - 13
Steven Austad, Judith Campisi, David Sinclair

**Nuclear Receptors & Disease**
October 30 - November 3
Ron Evans, Susanne Mandrup, Keith Yamamoto

**Personal Genomes & Medical Genomics**
November 14 - 17
Evan Eichler, Deanna Kroetz, Richard Lifton,
Jim Lupski, Richard Weinshilboum

**Neurodegenerative Diseases**
November 28 - December 1
Karen Duff, Jeffrey Rothstein, John Trojanowski

**Blood Brain Barrier**
December 5 - 8
Richard Daneman, Britta Engelhardt, Ryan Watts

## 2013 Spring Meetings Preview

**From Base Pair to Body Plan - Celebrating 60 Years of DNA**
February 28 - March 3
**Single Cell Analyses**
March 6 - 9
**Systems Biology: Networks**
March 13 - 16
**Computational Cell Biology**
March 19 - 22
**Cell Engineering & Cell-Based Therapies**
April 3 - 6
**RNA & Oligonucleotide Therapeutics**
April 10 - 13
**Synapses: From Molecules to Circuits & Behavior**
April 16 - 20
**Cancer Biology & Therapeutics**
April 23 - 27
**Telomeres & Telomerase**
April 30 - May 4
**The Biology of Genomes**
May 7 - 11
**The Ubiquitin Family**
May 14 - 18
**Retroviruses**
May 20 - 25
**78th Symposium: Immunity & Tolerance**
May 29 - June 3

## 2013 Early Spring Courses

**Workshop on Leadership in BioScience**
February 22 - 25
**Protein Purification & Characterization**
April 3 - 16
**Quantitative Imaging: From Cells to Molecules**
April 3 - 16
**Cell & Developmental Biology of *Xenopus***
April 5 - 16

Cold Spring Harbor Laboratory Meetings & Courses Program 1 Bungtown Road Cold Spring Harbor, NY 11724
516 367 8346 meetings@cshl.edu www.cshl.edu/meetings

NW227533E

a**nature**conference

*Nature Genetics* and the Wellcome Trust present:

## The Genomics of Common Diseases 2012

*Nature Genetics* in association with the Wellcome Trust is pleased to announce the sixth annual Genomics of Common Diseases meeting, which will address the following topics across a range of common diseases: exome and whole genome sequencing, clinical and targeted sequencing, regulatory and functional genomics, cancer genomics, clinical translation and pharmacogenetics, statistical analyses of sequencing datasets, population and evolutionary genetics and infectious disease and host genetics.

### September 19-22, 2012
### Bolger Center, Potomac, MD, USA

**KEYNOTE SPEAKERS**

**Francis Collins** (National Institutes of Health, USA)
**Thomas Hudson** (Ontario Institute for Cancer Research, Canada)
**Neil Risch** (University of California, San Francisco, USA)
**Sarah Tishkoff** (University of Pennsylvania School of Medicine, USA)

**SPEAKERS**

**Antonis Antoniou**
**John Blangero**
**Mary Carrington**
**Stephen Chanock**
**Paul de Bakker**
**Manolis Kellis**
**Matthew Meyerson**
**Patricia Munroe**
**Benjamin Neale**
**Christopher O'Donnell**
**Alkes Price**
**Jonathan Pritchard**
**Dan Roden**
**Nicholas Schork**
**Christine Seidman**
**Jay Shendure**
**Adam Siepel**
**Pamela Sklar**
**Nicole Soranzo**
**Shamil Sunyaev**
**Peter Visscher**
**George Weinstock**
**Cisca Wijmenga**
**Ramnik Xavier**

**ORGANIZERS:**

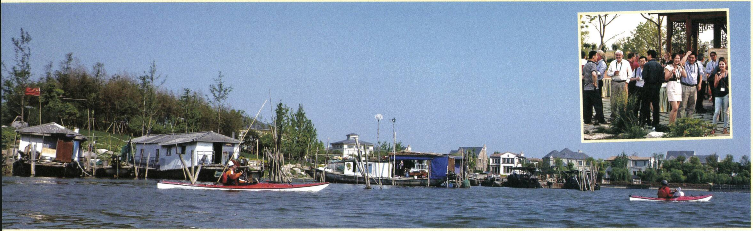
**Orli Bahcall** (*Nature Genetics,* USA)
**Aravinda Chakravarti** (Johns Hopkins University School of Medicine, USA)
**Peter Donnelly** (University of Oxford, UK)
**Michael Dunn** (Wellcome Trust, UK)
**Michel Georges** (University of Liège, Belgium)
**Teri Manolio** (National Human Genome Research Institute, National Institutes of Health, USA)
**Mark McCarthy** (University of Oxford, UK)
**Nazneen Rahman** (Institute of Cancer Research, UK)
**Mary Relling** (St. Jude Children's Research Hospital, USA)

For more information and to register visit:
**www.nature.com/natureconferences/gcd2012**

nature genetics

wellcome trust

IN222279



# Cold Spring Harbor Asia 2012 Fall Meetings

**Ion Channels: Biophysics, Diseases and Therapeutics**
September 17 - 21
Hualiang Jiang, Henry Lester, Min Li, Jianpeng Ma

**Small GTPases at Different Scales: Proteins, Membranes and Cells**
September 24 - 28
Jianping Ding, Klaus Gerwert,
Roger Goody, Shuh Narumiya

**RNA Biology**
October 8 - 12
Fátima Gebauer, Narry Kim,
Adrian Krainer, Mutsuhito Ohno

**High Throughput Biology: Genomics & Epigenomics**
October 15 - 19
Edward Rubin, Bing Ren, Tian Xu

**Differentiation Therapy and Advances in Cancer**
October 20 - 24
Saijuan Chen, Zhu Chen, Samuel Waxman

**Plant Epigenetics, Stress and Evolution**
October 29 - November 2
Jian-Kang Zhu, Yijun Qi, Brandon Gaut, Zhizhong Gong

**Neural Circuit Basis of Behavior and its Disorders**
November 5 - 9
Guoping Feng, Yasunori Hayashi,
Minmin Luo, Richard Morris

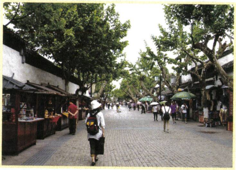
**CSH Asia / ICMS Conference on Tumor Microenvironment**
November 13 - 17
Frances Balkwill, Wolf Fridman, Luyuan Li,
Abraham Raz, Eiichi Tahara, Isaac Witz

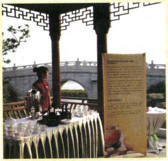
**Synthetic Biology**
November 26 - 30
Guoqiang Chen, Sang Yup Lee, Hiroki Ueda

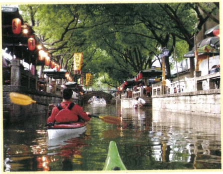
**Stem Cells and Developmental Mechanisms**
December 3 - 7
Thomas Bosch, Yukiko Gotoh, Ting Xie, Weimin Zhong

**Vaccine Design**
December 10 - 14
Michael Good, Hiroshi Kiyono, Bali Pulendran,
Yiming Shao, Linqi Zhang

**Cold Spring Harbor Asia**
Suzhou Dushu Lake Conference Center
Suzhou, China
**www.csh-asia.org**
meetings@csh-asia.org







NW227535E

# IF ONLY ...

*A taste of your own medicine.*

**BY TONY BALLANTYNE**

"Doctor," said Sacha, "Can you give me your assurance that this injection won't harm my children?"

"Well, there's always some risk, Ms Melham. I do have a leaflet that explains everything …"

Sacha placed a finger on the table.

"I don't need a leaflet, Doctor. I simply want your assurance that this injection will cause Willow and Gregory no harm …"

Doctor James Ferriday gazed at the finger.

"As I said, there is always a small risk, but if you look, you will see that this is less than the probability of …"

Sacha held up her hand.

"Please, Doctor. Don't try and confuse the issue."

"I'm not trying to confuse the issue, I'm simply presenting you with the facts …"

Sacha rose to her feet.

"Well, I think I've heard enough. Willow, Gregory, put your coats back on. Thank you, Doctor, we'll be … what's that?"

James's screen flashed red and green.

"Oh dear," he said, reading the yellow writing scrolling across the monitor. "I think you should take a seat."

Sacha did so. Her son slipped his hand into hers.

"What's the matter, mummy?"

"Nothing, dear. Is everything OK, Doctor?"

"I'm sorry, Ms Melham …" he began, and then more kindly. "I'm sorry, Sacha, but you've crossed the threshold. I'm afraid to say, you're not allowed science any more."

"I'm what?"

"You're not allowed science any more," repeated James.

Sacha's lips moved as she tried to process what he had said.

"You're saying that you're refusing my children treatment?"

"No," said James. "Quite the opposite. You and your children will always be entitled to the best medical care. It's just that you, Sacha, no longer have a say in it. I shall administer the vaccination immediately."

"What?" Sacha sat up, eyes burning with indignation. "How dare you? I, and my husband, are the only ones who say how my family is run."

"Well, yes," said James. "But you no longer have a say in things where science is involved. You're not allowed science any more."

"I never heard anything so ridiculous! Who decided that?"

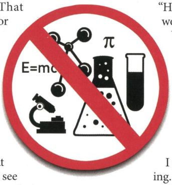
"The Universe."

"The Universe? Why should the Universe say I'm not allowed science any more?"

"Because you haven't paid science enough attention. You've had the opportunity to read the facts and the education to be able to analyse them, yet you have consistently chosen not to."

"The education?" exclaimed Sacha. "Hah! My science education was terrible. None of my teachers could explain anything properly."

"Really?" said James. "That would certainly be grounds for appeal …"



JACEY

He pressed a couple of buttons. Tables of figures appeared on the screen.

"No," he said, shaking his head. "I'm sorry … it turns out that your teachers were all really rather excellent. You went to a very good public school, after all. If you look at your teachers' results you will see they added significant value to their pupils' attainment."

Sacha pouted.

"Well, they didn't like me."

"Possibly …"

He pressed a couple more buttons.

"What?" said Sacha, hearing his sharp intake of breath.

"Look at this," said James, scrolling down a long table. "Times and dates of occasions when you've proudly admitted to not being good at maths."

"What's the matter with that? I'm not."

"It's not the lack of ability, Sacha, it's the fact that you're proud of it. You'd never be proud of being illiterate. Why do you think your innumeracy is a cause for celebration?"

"Because … Well …"

"That's why you're not allowed science any more."

"This is outrageous!" snarled Sacha. "How can this happen?"

"Oh, that's easy," said James. "Magic."

"Magic?" said Sacha, her eyes suddenly shining. "You mean there's really such a thing?"

"Of course not. But I can't explain to you how it's really done because you're not allowed science any more."

Sacha fumbled for her handbag.

**➲ NATURE.COM**
**Follow Futures on Facebook at:**
**go.nature.com/mtoodm**

"I'm calling the BBC," she said. "I'm a producer there, you know. I'll report you."

"Report me to who you like," said James. "The story will never get out. All your cameras and microphones and things work on science."

Sacha gazed at him.

"Who gave you the right to control my life?"

"You've got it the wrong way round. You gave the right to control your life away. You're the one who chose to ignore the way the world works."

"Hah!" said Sacha. "The way the world works! Bloody scientists. You think the world is all numbers and machines and levers. You don't understand anything about the soul or spirit."

"Of course I do," said James. "I've been happily married for 20 years. I have two children that I love. I play the piano, I enjoy reading. It's just that I have additional ways of looking at things."

Sacha stood up.

"Willow, Gregory. We're going home," she glared at James. "That is if I'm still allowed to drive? You don't have something against women drivers as well do you, Doctor?"

"This is nothing to do with you being female, Ms Melham," said James, calmly. "This is purely about your attitude to science. Now, before you go, I'll administer the injection to the three of you."

"You will not! I will not allow it."

"I told you, you have no choice."

"Why? Because I disagree with you?"

For this first time, James's anger showed itself.

"No!" he snapped. "You don't get it! You're allowed to disagree with me, I want you to disagree with me! I'd love to engage in reasoned debate with you. But until you take the trouble to understand what you're talking about, you're not allowed science any more. Now, roll up your sleeve."

Sacha muttered something under her breath.

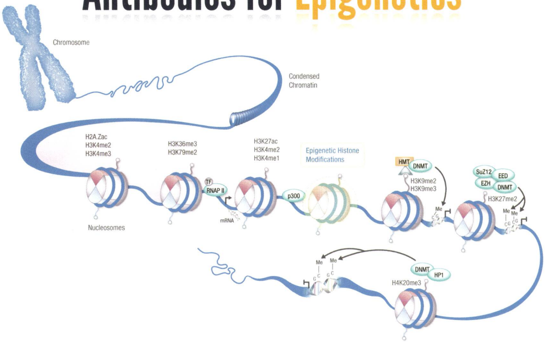
"What's in the injection?" said James. "You know, you start asking questions like that, you might get science back …" ■

**Tony Ballantyne**'s *latest collection of tales is* Stories of the Northern Road *(NewCon Press). You can find him at tonyballantyne.wordpress.com.*

**EPITOMICS®**
an **abcam®** company

*Now Available*
**3500 RABBIT MONOCLONALS**
*Over 7500 Antibodies Total*

# Antibodies for Epigenetics



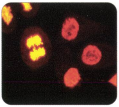
Chromosome
Condensed Chromatin
H2A.Zac
H3K4me2
H3K4me3
H3K36me3
H3K79me2
H3K27ac
H3K4me2
H3K4me1
Epigenetic Histone Modifications
HMT
DNMT
H3K9me2
H3K9me3
SuZ12
EED
EZH
DNMT
H3K27me2
Me
Me
Me
TF
RNAP II
mRNA
p300
Nucleosomes
Me
Me
DNMT
HP1
H4K20me3

**Epigenetics**

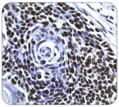
Our Rabbit Monoclonal Antibodies (RabMAbs®) combine the benefits of the high affinity from the rabbit, with the specificity of a monoclonal, enabling the development of high quality modification-specific antibodies (1, 2). We have now generated more than 500 modification-specific and activation-specific antibodies using our patented technology, including our new 5-Methylcytosine RabMAb (Cat # 3913-1, Clone EP4694) which can be used to identify methylated DNA to study methylation events.

Additionally, we validate every RabMAb in WB, IHC, IF, IP and FACS before release, so you can be assured of which assays the antibody will work in before purchasing.

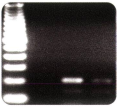
## High Quality Antibodies for Epigenetics



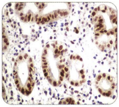
**Histone H3 (pS10)**
Cat.#: 1173-1



**Histone H4 (K8)**
Cat.#: 1796-1



**5-Methylcytosine**
Cat.#: 3913-1



**Dnmt1**
Cat.#: 3147-1

## Learn More @ www.epitomics.com/epigenetics

1. Rabbit monoclonal antibodies: a comparative study between a novel category of immunoreagents and the corresponding mouse monoclonal antibodies, Rossi S. et al., Am J Clin Pathol. 2005 Aug;124(2):295-302
2. Rabbit monoclonal antibody: potential application in cancer therapy, Feng L. et a;. Am J Transl Res. 2011 May 15; 3(3): 269–274

100% SATISFACTION GUARANTEE *Antibody Satisfaction Guarantee*

**www.epitomics.com**

US & Canada | 1-877-772-2622
Outside NA | 650-583-6688
info@epitomics.com

**AMPLIFICATION** // QX100™ DROPLET DIGITAL™ PCR SYSTEM

The future of qPCR is here, and it's digital.

BIO-RAD
QX100™ Droplet Reader
BIO-RAD
QX100™ Droplet Generator

That's **ddPCRevolutionary.**

Bio-Rad's QX100™ Droplet Digital™ PCR system provides a measure of target DNA molecules with unrivaled precision and accuracy. The QX100 system partitions each sample into 20,000 individual nanoliter-sized droplets. PCR-positive and PCR-negative droplets from every sample are then counted to provide absolute target quantification in digital form. The QX100 system provides a revolutionary approach to target DNA quantification.

- Detect rare target sequences with unmatched sensitivity for cancer and viral research
- Measure small differences in target copy number variations (down to 1.2x differences)
- Determine gene expression levels without a standard curve or ΔΔCq

Visit **www.bio-rad.com/ad/digitaldroplet3** or contact your Bio-Rad Sales Representative to learn more.

**Research. Together.**

To find your local sales office, visit **www.bio-rad.com/contact**.
In the U.S., call toll free at 1-800-4BIORAD (1-800-424-6723).

Visit us at www.bio-rad.com

BIO-RAD